

Consortium HarmoS Langue étrangères

LANGUE ÉTRANGÈRES

Rapport scientifique de synthèse et modèle de compétences

Version provisoire (avant adoption des standards de base)

Etat : 18 juin 2009

Consortium scientifique :

Université de Fribourg (Leading house) | Günther Schneider, Peter Lenz, Thomas Studer
(rédacteurs principaux du rapport)

Pädagogische Hochschule Zürich | Sandra Hutterli, Ruth Keller

Universität Bern | Verena Tunger

Institut de Recherche et de Documentation | Claudine Brohy

Pédagogique (IRD), Neuchâtel | Daniel Elmiger

Haute Ecole pédagogique de Fribourg | Christine Müller

Pädagogische Hochschule Zentralschweiz Luzern | Monika Mettler

Table des matières

1	Introduction	3
1.1	Remarque préliminaire	3
1.2	Mandat	3
1.3	Consortium – composition et exécution du mandat	3
2	Bases de travail	5
2.1	Modèle de compétences	5
2.2	Vérification du «construct» (méthodologie).....	11
3	Les enquêtes empiriques menées dans le cadre d’HarmoS.....	15
3.1	Vue d’ensemble	15
3.2	Ampleur des enquêtes	15
3.3	Instruments de test.....	18
3.3.1	Réflexions fondamentales sur l’évaluation des compétences	18
3.3.2	Elaboration des tests dans HarmoS	21
3.4	Compréhension de l’oral	24
3.5	Compréhension de l’écrit.....	27
3.5.1	«Construct» et bases conceptuelles.....	27
3.5.2	Instruments	32
3.5.3	Collecte des données.....	38
3.5.4	Correction	38
3.5.5	Résultats et analyse des résultats	38
3.5.6	Bilan et perspectives	46
3.5.7	Résumé	48
3.6	Ecrire	49
3.6.1	Situation initiale.....	49
3.6.2	Instruments (exercices).....	49
3.6.3	Collecte des données.....	53
3.6.4	Evaluation	53
3.6.5	Résultats de l’enquête.....	56
3.6.6	Analyses des résultats	61
3.6.7	Etat des travaux et perspectives	65
3.7	Parler	65
3.7.1	Situation initiale et bases de travail	65
3.7.2	Instruments (exercices).....	67
3.7.3	Collecte des données.....	72
3.7.4	Evaluation	72
3.7.5	Analyses et résultats	74
3.7.6	Perspectives: examen de la compétence orale dans le cadre du monitoring de l’éducation.....	77
4	Domaines de compétences n’ayant pas été étudiés empiriquement	78
4.1	Médiation linguistique.....	78

4.2	Compétences transversales	81
4.2.1	Généralités.....	81
4.2.2	Compétences culturelles et interculturelles	83
4.2.3	Compétences méthodologiques	90
4.2.4	Références bibliographiques relatives aux compétences transversales	95
5	Propositions des experts pour des standards de base	100
6	Références bibliographiques.....	102

1 Introduction

1.1 Remarque préliminaire

Le présent document propose une vue d'ensemble du projet «HarmoS - Langues étrangères». Dans ce rapport, qui adopte une approche plutôt scientifique, l'accent est mis sur les enquêtes empiriques qui ont été réalisées dans le cadre du projet sur les domaines de compétences suivants: «Compréhension de l'écrit», «Ecrire» et «Parler» (chap. 3); le rapport comporte également des réflexions sur la médiation linguistique et sur certaines compétences transversales (chap. 4). Les chapitres mentionnés sont précédés d'une esquisse du contexte dans lequel s'inscrit le projet (chap. 1), ainsi que d'une brève présentation des bases sur lesquelles se fondent les enquêtes (modèle de compétences et méthodologie, chap. 2). Enfin, le rapport expose la procédure qui a été suivie pour fixer les standards de formation de base (chap. 5).

1.2 Mandat

La CDIP a confié au consortium HarmoS Langues étrangères (LÉ) le mandat d'élaborer, pour les principales disciplines «langues étrangères» de la scolarité obligatoire (à savoir l'allemand et l'anglais en Suisse romande; le français et l'anglais en Suisse alémanique), des propositions d'experts concernant des standards de formation qui se rapportent à des compétences de base devant être acquises au terme du degré primaire d'une part et du degré secondaire I d'autre part. Ces propositions doivent reposer sur un modèle de compétences validé d'un point de vue didactique mais aussi, dans la mesure du possible, par voie empirique. La validation empirique devait aussi être mise à profit pour obtenir des points de repère fiables sur ce que les élèves qui se trouvent actuellement à ces périodes de transition (donc des élèves de sixième et de neuvième années) sont capables de fournir en termes de performances.

La situation du consortium Langues étrangères est particulière dans le sens où des standards de base doivent être fixés pour de futurs élèves, qui auront appris les langues étrangères dans des conditions différentes. En effet, les nouveaux scénarios pour l'enseignement des langues étrangères, lesquels ont été définis dans la décision du 25 mars 2004 de la Conférence des directeurs de l'instruction publique, n'ont, pour la majeure partie, pas encore été mis en œuvre. Cette décision règle notamment le début de l'enseignement des langues étrangères: l'enseignement de la première langue étrangère doit commencer au plus tard durant la troisième année actuelle, et la deuxième langue étrangère doit être introduite en cinquième année au plus tard. L'une de ces langues doit être une langue nationale de la Suisse; l'autre langue introduite est l'anglais. Au terme de leur scolarité obligatoire, les élèves doivent avoir une maîtrise comparable des deux langues étrangères. Parallèlement à l'introduction de deux langues étrangères au degré primaire, il est aussi prévu de renouveler l'enseignement dans ces disciplines. De nouveaux plans d'études et supports d'enseignement sont développés dans ce but¹. Pour l'instant, il n'est donc guère possible d'établir plus que des hypothèses quant aux répercussions qu'entraîneront ces changements – quelles seront, aux périodes de transition de la scolarité obligatoire, les performances en langues étrangères des élèves qui auront connu les nouvelles conditions d'enseignement? Les suppositions avancées sur ces questions se basent sur des rapports de recherche, sur le consensus des experts et sur des extrapolations de résultats obtenus dans le cadre de l'enseignement des langues étrangères, tel qu'il est dispensé actuellement.

1.3 Consortium – composition et exécution du mandat

La CDIP a chargé le Centre d'Enseignement et de Recherche en Langues Etrangères (CERLE) de l'Université de Fribourg de diriger le projet en tant que mandataire; c'est le CERLE qui assume la fonction de *leading house* pour le consortium Langues étrangères (responsabilité du projet: Günther Schneider; direction de projet: Peter Lenz, Thomas Studer; secrétariat: Eva

¹ Se référer à Caspari 2005, notamment, pour les différents objectifs poursuivis et l'interaction visée entre les standards, les plans d'études et les moyens d'enseignement.

Wiedenkeller). Ce projet a été mené avec les partenaires suivants (dans l'ordre de leur participation): la *Pädagogische Hochschule Zürich* (Sandra Hutterli, Ruth Keller), l'*Université de Berne* (Verena Tunger), l'Institut de recherche et de documentation pédagogique (IRDP) de Neuchâtel (Daniel Elmiger, avec la contribution de Claudine Brohy), la Haute Ecole pédagogique de Fribourg (Christine Müller-Tragin), ainsi que la *Pädagogische Hochschule Zentralschweiz* de Lucerne (Monika Mettler). La *Pädagogische Fachhochschule Graubünden* de Coire (Luigi Menghini) a également contribué au projet avec un statut d'observateur. Le travail réalisé dans le cadre de ce projet n'a été que partiellement financé par le budget de projet alloué par la CDIP; les institutions mentionnées ont dû elles aussi réunir des fonds considérables à cet effet.

Le consortium a été accompagné par un conseil consultatif, qui a été sollicité à plusieurs reprises pour des décisions centrales, notamment lors de la phase de définition des exigences que les standards de base devraient poser. Ce groupe était constitué des membres suivants: Bruno Dahinden, canton de Thurgovie; Mirjam Egli, CDIP et Uni BS; Simona Pekarek, Uni NE; Michel Nicolet, secrétariat de la CIIP; Véronique Roncoroni-Arletta, puis Brigitte Jörimann Vancheri, canton du Tessin; Victor Saudan, responsable des langues à la NW EDK; Daniel Stotz, HEP ZH; Iwar Werlen, Uni BE; Susanne Wokusch, HEP VD et UNIL.

Par ailleurs, l'aide d'autres personnes a également été sollicitée pour la réalisation des tests d'expression orale (production et interaction), ainsi que pour la correction et pour l'évaluation des performances des élèves. Quelques didacticiennes et didacticiens spécialisés, qui ne faisaient partie ni du consortium ni du conseil, ont aussi apporté leur contribution, en tant qu'experts et partenaires de discussion, pour l'élaboration des exercices des tests, de même que pour la définition des standards de base. Les tests préliminaires réalisés avec les exercices ainsi conçus et l'exécution des tests menés sur le terrain dans le domaine de l'expression orale n'auraient pas été possibles sans la collaboration de nombreux enseignants et enseignantes, ainsi que de leurs directions d'école et de leurs classes.

Les différents collaborateurs et collaboratrices du consortium ont principalement fixé leurs priorités de travail en fonction des domaines de compétences qui devaient faire l'objet d'un examen et d'une description (c'est-à-dire surtout: lire, écrire, parler). Il n'y a donc pas eu par exemple des sous-groupes régionaux ou des groupes qui se seraient exclusivement concentrés sur l'une des trois langues cibles. La direction du consortium a également été active au sein du groupe de la CDIP chargé de la stratégie et a pris ses décisions concernant l'enquête principale en collaboration avec le groupe HarmoS chargé des questions de méthodologie. Les membres de ce groupe ont été principalement actifs dans l'analyse statistique des données. La collaboration s'est bien passée dans l'ensemble. Durant la phase initiale, un aspect s'est révélé négatif, à savoir que le groupe «Méthodologie» d'HarmoS n'a pu intervenir que trop tard, puisqu'il a aussi d'abord dû se constituer. Quelque temps avant le délai de livraison officiel pour les rapports des consortiums (décembre 2007), il s'est avéré qu'il ne restait pas suffisamment de temps pour réaliser différentes analyses en rapport avec l'exploitation de l'enquête principale, du fait que le groupe «Méthodologie» devait apporter son soutien à tous les consortiums en parallèle. Durant l'année 2008, le consortium Langues étrangères a ensuite poursuivi le travail en assumant une partie de ces analyses. L'idée de la direction de projet HarmoS, formulée au printemps 2006, de réaliser l'enquête principale une année plus tard – pour ainsi permettre que davantage d'enquêtes approfondies soient réalisées – quoique prometteuse au départ, a finalement entraîné une perte de temps considérable pour le consortium Langues étrangères. Ce dernier a alors décidé d'ajourner les travaux restants et d'élaborer une nouvelle planification, laquelle n'a cependant pas été acceptée, à l'automne 2006, par les Directeurs de l'instruction publique. Ce refus était regrettable, non seulement pour des questions de planification, mais aussi du point de vue du fond; en effet, des enquêtes qualitatives complémentaires sur les tests – enquêtes auxquelles il a fallu renoncer à la suite du refus de la nouvelle planification – auraient certainement pu apporter des éléments essentiels pour la validation du modèle de compétences et des tests qui devront désormais permettre d'évaluer les standards (cf. Banerjee 2004). Mais l'argument déterminant pour le mandant était visiblement qu'une enquête représentative aujourd'hui ne pourrait, même avec un concept de validation pointu, fournir des points de repère exacts quant aux compétences des futurs élèves, qui suivront leur scolarité dans les conditions qu'offriront les nouveaux scénarios.

Du fait de cette position, mais aussi en raison du manque de temps qui s'est fait sentir, les langues étrangères sont représentées par deux domaines de compétences uniquement dans les enquêtes représentatives principales en Suisse alémanique et en Suisse romande: il s'agit de la compréhension de l'écrit et de la capacité à écrire. Cependant, le Secrétariat général de la CDIP a évoqué la possibilité de lancer une enquête supplémentaire plus englobante une fois que les nouveaux scénarios d'enseignement auront déployé leurs effets.

2 Bases de travail

2.1 Modèle de compétences

Le mandat de la CDIP confié aux consortiums prévoit que le modèle de compétences et les standards puissent être mis en relation. Il s'agit de développer et de valider empiriquement (du moins en partie) un modèle de compétences sur la base duquel il est possible de définir et de communiquer les exigences de base pouvant être posées aux élèves de différents degrés.

Les attentes de la CDIP concernant ce modèle de compétences sont essentiellement liées à l'expertise (Klieme et al. 2003)² qui a été réalisée pour le compte de la Conférence allemande des ministres de l'éducation et des affaires culturelles des länder (KMK), afin de servir de document de travail pour le développement de standards de formation. Dans cette étude, la notion de "compétence" est définie en ces termes – en référence à Weinert:

En accord avec Weinert (2001, p. 27 sqq.), nous entendons par compétences «les capacités et aptitudes cognitives dont l'individu dispose ou qu'il peut acquérir pour résoudre des problèmes précis, ainsi que les dispositions motivationnelles, volitives et sociales qui s'y rattachent pour utiliser avec succès et responsabilité les résolutions de problèmes dans des situations variables». (Klieme et al. 2003: 74 [en gras dans l'original])

La compétence se comprend ainsi essentiellement comme une capacité (cognitive) à résoudre des problèmes pouvant être mise en œuvre en interaction étroite avec certaines «dispositions» non cognitives. La définition de la compétence à laquelle se réfèrent Klieme et al. est très proche du modèle du *Cadre européen commun de référence pour les langues* (Conseil de l'Europe 2001), sur lequel le consortium Langues étrangères se base principalement. D'ailleurs, dans l'annexe de l'étude Klieme (Klieme et al. 2003: 149-159), le modèle du *Cadre de référence* et ses descriptions de compétences (descripteurs) sont présentés comme «une bonne base pour le développement d'exercices, de tests et d'autres procédés d'examen».

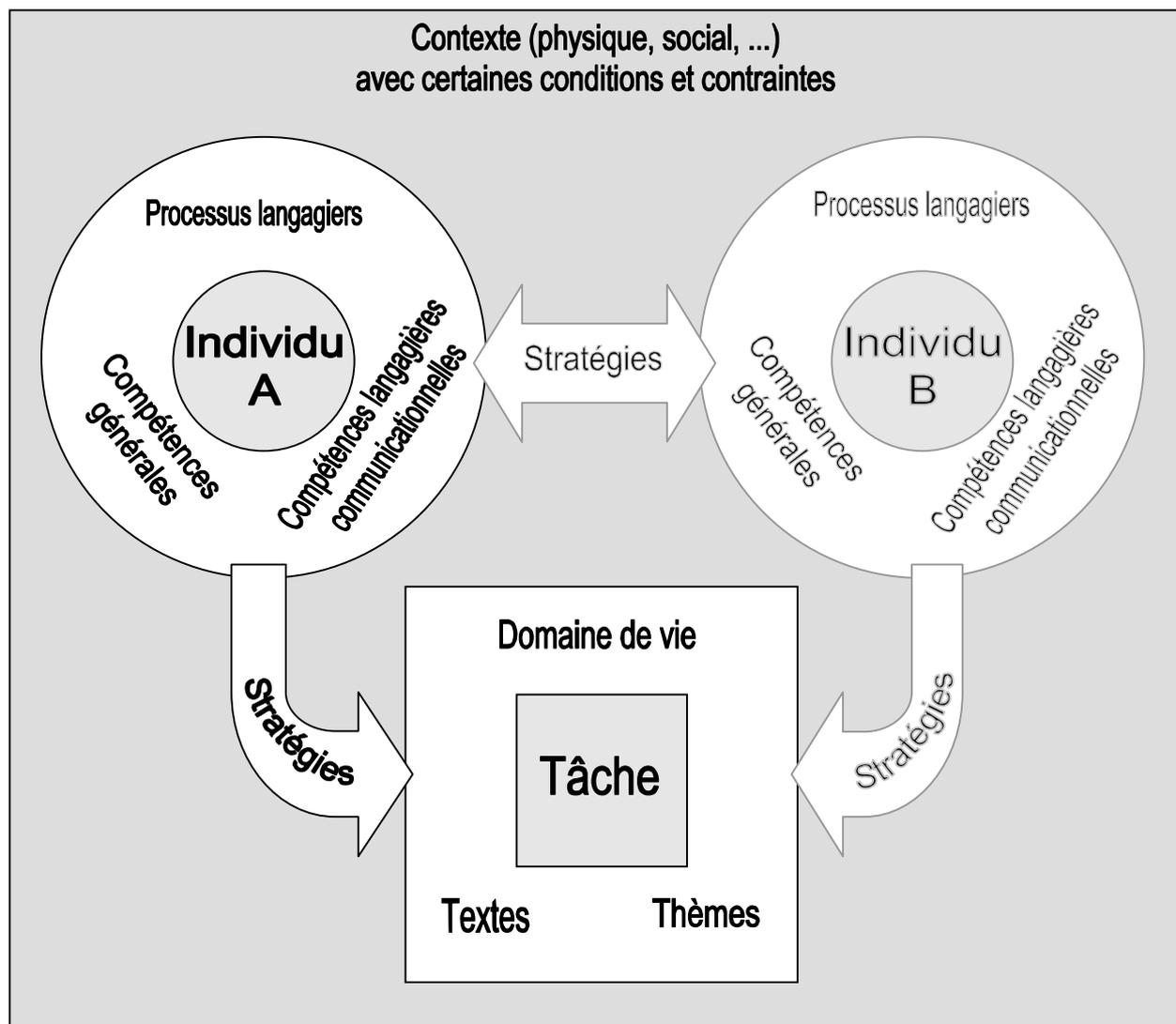
C'est pour le consortium Langues étrangères un avantage certain que de pouvoir se référer dans ses travaux de développement à un modèle de compétences déjà existant, qui a été spécifié, de manière parfois très complète, non seulement dans le *Cadre européen commun de référence*, mais aussi dans différents projets basés sur le cadre; de plus, ce modèle est très bien accepté dans différents milieux. Dans le domaine de la scolarité, ce modèle de compétences s'est fait connaître (dans les grandes lignes) au travers du *Portfolio européen des langues* surtout, mais aussi par le biais de certains plans d'études cantonaux ou régionaux assez récents.

Le modèle du *Cadre de référence* (ou du moins certaines parties du cadre) est déjà si largement répandu dans la didactique des langues étrangères – et particulièrement dans le domaine des tests de langues (et des diplômes de langues connus) – qu'il semble absolument naturel de le prendre pour point de départ des travaux dans le cadre du projet HarmoS; il serait tout à fait illogique de renoncer à y faire référence. En tant que mandant, la conférence des directeurs de l'instruction publique attend également des standards HarmoS qu'ils se rapportent explicitement au *Cadre de référence* ainsi qu'aux compléments apportés par *IEF / lingualevel* (Lenz / Studer 2007) et aux différenciations établies dans ce cadre par groupe cible (tranche d'âge des 11 à 16 ans). Sans compter que la *leading house* a contribué de manière substantielle à l'élaboration des descriptions de compétences afférentes aux niveaux de référence et qu'elle a participé à certains travaux importants réalisés par la suite, sur la base du *Cadre de référence* (notamment *IEF / lingualevel*, mais aussi des évaluations de performances de référence); elle adhère donc au modèle de base évoqué.

² Il est aussi possible dorénavant de se référer à l'expertise détaillée d'Oelkers / Reusser (2008), qui n'a toutefois plus pu être prise en compte dans le cadre du présent rapport.

Dans le *Cadre de référence*, la compétence langagière est perçue en premier lieu comme une compétence à utiliser la langue, donc comme la capacité d'agir en ayant recours à une langue et ainsi de réussir à accomplir des tâches de diverse nature. Si la langue est donc centrale dans ce modèle, elle ne l'est pas dans un sens réducteur: la compétence à utiliser une langue est décrite comme une capacité humaine globale, qui engage l'être humain «tout entier», avec toutes ses ressources linguistiques et non linguistiques, qu'elles soient d'ordre cognitif ou autre, par ex. affectives.

Le schéma suivant donne un aperçu du modèle d'utilisation de la langue dans le *Cadre européen de référence*:



III. 1: Utilisation de la langue selon le *Cadre européen de référence*

Le schéma figurant ci-dessus se lit ainsi: l'utilisation de la langue à des fins de communication se produit dans les conditions spécifiques d'un contexte. Les individus communicants disposent de ressources composées de toutes sortes de compétences, qui se répartissent en compétences générales et en compétences langagières communicationnelles spécifiques. Lorsqu'une personne produit ou reçoit un acte langagier, des processus cognitifs se déclenchent, qui permettent de traiter spécifiquement l'information linguistique (processus langagiers). Des stratégies sont déployées pour piloter les activités langagières – interaction avec d'autres ainsi que production et réception de textes écrits ou oraux – ce qui se fait plus ou moins consciemment. Les tâches de communication langagières ont certaines caractéristiques, qui permettent de les différencier grosso modo: le domaine de vie dans lequel elles s'inscrivent – formation (école), vie publique,

etc. – les textes oraux ou écrits qui doivent être traités en rapport avec la tâche, les thèmes dont traitent ces textes. Dans les activités langagières interactives, une partie essentielle de la tâche consiste à interagir avec d'autres individus (par ex. dans des discussions nécessaires pour prendre une décision; dans la correspondance écrite). Toutefois, de nombreuses activités langagières de réception et de production se conçoivent typiquement comme des processus qui se déroulent entre un individu et une tâche. Les autres individus jouent alors le rôle de contexte social et ne sont qu'indirectement concernés par le traitement des tâches (voir illustration 1, individu en gris)³.

Le *Cadre de référence* détaille, sous différents angles et de manière plus ou moins approfondie, les compétences qui composent la «compétence à communiquer langagièrément» ou qui y contribuent.

- 1) La description de la compétence, dans le *Cadre de référence*, se concentre sur la caractérisation de la **capacité de communication** à différents niveaux, et ce au moyen de descriptions d'«activités de communication langagière» et de «stratégies» (chap. 4.4), pour la plupart étalonnées⁴ de façon empirique (c'est-à-dire classées selon des niveaux de compétence). Elles illustrent ce qu'une personne d'un certain niveau sait en principe faire, quelles sont les *activités langagières* qu'elle maîtrise à ce stade. Le *Cadre de référence* répartit les activités et les stratégies de communication langagière selon quatre modes d'action langagière différents (réception, interaction, production et médiation⁵), avec à chaque fois une distinction entre l'usage oral de la langue et son usage écrit (canal).

		Mode de communication			
		Production	Interaction	Réception	Médiation
Canal	oral	■	■	■	■ ■ ■
	écrit	■	■	■	■ ■ ■

III. 2: Système de classement des activités langagières selon le mode de communication et le canal utilisés (en référence au *Cadre européen de référence*)

Dans ce tableau, la combinaison «réception» et «écrit» correspond, par exemple, à cette aptitude langagière de communication qui est la compréhension de l'écrit. En tout, le *Cadre de référence* compte plus de 50 échelles pour les activités et les stratégies de communication langagière.

- 2) Le *Cadre de référence* contient également une description des **ressources qui permettent la mise en œuvre de compétences** (chap. 5) et qui s'avèrent nécessaires pour réaliser des activités langagières. On y établit une distinction entre les compétences générales et les compétences spécifiquement langagières:

³ Dans le *Cadre de référence* même, on trouve la définition suivante, qui décrit très succinctement les caractéristiques du modèle d'utilisation de la langue, donc dans une perspective de type actionnel: «L'usage d'une langue, y compris son apprentissage, comprend les actions accomplies par des gens qui, comme individus et comme acteurs sociaux, développent un ensemble de **compétences générales** et, notamment une **compétence à communiquer langagièrément**. Ils mettent en œuvre les compétences dont ils disposent dans des **contextes** et des **conditions** variés et en se pliant à différentes **contraintes** afin de réaliser des **activités langagières** permettant de traiter (en réception et en production) des **textes** portant sur des thèmes à l'intérieur de **domaines** particuliers, en mobilisant les **stratégies** qui paraissent le mieux convenir à l'accomplissement des **tâches** à effectuer. Le contrôle de ces activités par les interlocuteurs conduit au renforcement ou à la modification des compétences» (Conseil de l'Europe, 2001:15)

⁴Ces travaux de développement ont été réalisés principalement dans le cadre du programme national de recherche sur l'«Efficacité de nos systèmes de formation». La responsabilité du projet avait été confiée à ce qui est devenu le Centre d'Enseignement et de Recherche en Langues Etrangères de l'Université de Fribourg (CERLE – cf. par ex. North & Schneider 1998; Schneider & North 2000; North 2000). (Entre-temps, le CERLE a également évolué; la nouvelle structure de l'Université de Fribourg porte l'appellation de «Département des Sciences du plurilinguisme et des langues étrangères».)

⁵ Fondamentalement, les activités langagières de médiation consistent à retransmettre, oralement ou par écrit, pour une personne qui ne comprend pas la langue utilisée, le sens de communications que l'on a entendues ou lues, dans une autre langue qu'elle comprendra. Il n'est donc pas question d'exprimer ses réflexions et ses intentions personnelles.

a) Compétences générales:

- Savoir: connaissance du monde (par ex. histoire, géographie); savoir socioculturel (par ex. sur la vie quotidienne; la vie culturelle; les valeurs; les relations interpersonnelles; le langage du corps); conscience interculturelle (par ex. en ce qui concerne les similitudes et les différences qui existent entre des communautés);
- Savoir-faire: aptitudes pratiques (par ex. comportement conforme aux conventions de la société; capacités professionnelles); aptitudes interculturelles (par ex. savoir établir une relation entre deux cultures);
- Savoir-être (compétence liée à la personnalité): par ex. la motivation; le style cognitif; les valeurs;
- Savoir-apprendre: par ex. conscience linguistique et communicationnelle; techniques d'apprentissage; capacité à gérer ce qui est nouveau.

b) Compétences communicatives langagières (au sens strict):

- Compétences linguistiques: par ex. lexicales; grammaticales; phonologiques;
- Compétences sociolinguistiques: par ex. les marqueurs linguistiques des relations sociales; les règles de politesse; faire usage de divers registres de langue; reconnaître différents dialectes et accents;
- Compétences pragmatiques: par ex. compétence discursive; compétence fonctionnelle; aptitude à utiliser différents schémas d'interaction. Des éléments qualitatifs importants de la compétence pragmatique sont la fluidité et la précision de l'expression.

Ces compétences constituent des ressources qui interviennent, selon une combinaison et un degré variables, dans toutes les activités langagières. Inversement, toutes les expériences déclenchent, du fait de l'action linguistique qu'elles impliquent, des processus d'apprentissage, influençant ainsi à leur tour la portée et la qualité des ressources permettant la mise en œuvre de compétences.

Comme pour la capacité de gestion communicationnelle, le *Cadre de référence* définit aussi pour les compétences communicatives langagières une série d'échelles définies par voie empirique, correspondant au système de niveaux A1 à C2, dont des échelles évaluant notamment l'étendue du vocabulaire, la correction grammaticale, la souplesse et l'aisance.

- 3) Comme autre domaine de compétences du *Cadre de référence* – un domaine qui intervient en amont des ressources permettant la mise en œuvre de compétences – on peut nommer les **stratégies de pilotage actionnelles**. Ces dernières dirigent l'accès aux différentes ressources (compétences) et permettent ainsi au sujet d'accomplir des tâches de communication concrètes (stratégies pour la planification de l'action, l'exécution, le contrôle des résultats et, le cas échéant, la remédiation dans l'action; cf. à ce sujet Bachman & Palmer 1996: 63).

Le système descriptif du *Cadre de référence* tient expressément compte du fait que le développement et la mobilisation de la compétence de communication n'impliquent pas uniquement des dimensions linguistiques, mais par exemple aussi une conscience socioculturelle, des aptitudes interculturelles, l'expérience et l'imagination, la motivation, les relations affectives, le savoir-apprendre, etc. Les «compétences» sont définies comme «l'ensemble des connaissances, des habiletés et des dispositions qui permettent d'agir» (Conseil de l'Europe, 2001: 15).

Dans le *Cadre de référence*, il manque des échelles dans de nombreux domaines, comme sur le savoir socioculturel, sur les aptitudes interculturelles, sur les stratégies d'apprentissage, mais aussi sur certains domaines relevant de la capacité de communication: sur la médiation linguistique, sur la lecture de textes littéraires ou encore sur les conversations téléphoniques. Toutefois, cela ne signifie aucunement – comme cela est parfois avancé de manière un peu hâtive – que ces domaines ne sont pas importants pour le *Cadre de référence*.

Dans certains de ces domaines qui n'ont pas d'échelles propres, rien ne s'opposerait à ce que des échelles soient élaborées dans le sens des échelles qui existent pour les compétences langagières (par exemple dans la médiation linguistique ou d'autres domaines de la capacité de communication). D'autres éléments de compétences ne pourraient pas vraiment être mesurés sur

une échelle; du moins pas parallèlement aux dimensions de la capacité de communication (surtout les «compétences générales»).

Les compétences non mesurables du sujet communicant ne peuvent être dégagées du modèle dans le *Cadre de référence*, puisqu'elles interviennent dans chaque acte de communication et qu'elles sont déterminantes pour la réussite ou l'échec de la communication. De toute évidence, les compétences non mesurables s'avèrent aussi particulièrement importantes pour les activités extrascolaires et, par exemple, également pour la réussite de tests de langues.

Initialement, il était prévu de développer, dans le cadre du projet «HarmoS - Langues étrangères», le modèle de compétences dans les domaines de la médiation linguistique, des compétences (inter)culturelles ainsi que du plurilinguisme, parce qu'il s'agit de domaines qui ne cessent de gagner en importance pour l'enseignement des langues étrangères dans le cadre scolaire.

Pour ce qui est de la médiation linguistique, il a surtout été possible de fournir un travail d'ordre conceptuel. Cependant, les approches développées n'ont pu faire l'objet d'un examen empirique. Pour pouvoir réaliser des avancées substantielles dans les domaines de la compétence (inter)culturelle et du plurilinguisme, il aurait été nécessaire de mener d'autres projets séparément (par ex. des projets du Fonds national suisse, dans le cadre du PNR 56). Si de tels projets n'ont pas été mis en œuvre, il a toutefois été possible de réaliser des travaux préliminaires pour les compétences (inter)culturelle ainsi que méthodologique, travaux qui peuvent être mis à profit dans d'éventuels projets HarmoS subséquents, de même que dans le travail sur les plans d'études et les moyens d'enseignement (cf. chap. 4.2, «Compétences transversales»).

Les travaux menés dans le cadre d'HarmoS reposent non seulement sur le *Cadre de référence* lui-même, mais aussi sur des développements et des concrétisations découlant du *Cadre de référence*. Il s'agit principalement:

- des résultats du projet alémanique *IEF* (produit commercial *lingualevel*), surtout les descriptions de la capacité de communication, qui ont été adaptées aux besoins et aux possibilités des élèves entre 11 et 16 ans environ; avec en plus des critères complémentaires pour une description (plus affinée) des caractéristiques linguistiques des performances orales et écrites (Lenz / Studer 2007; 2004);
- de documents basés sur le *Cadre de référence*, qui le complètent et qui ont été réalisés dans le cadre de projets du Conseil de l'Europe, à l'instar du manuel pour la classification des examens de langues (Conseil de l'Europe 2003); des exemples classés et commentés d'exercices et de performances notamment en rapport avec la compréhension de l'écrit et de l'oral, ainsi qu'avec l'expression orale et écrite, etc.⁶;
- de travaux scientifiques portant sur le *Cadre de référence*;
- du *Portfolio européen des langues* et de travaux réalisés sur le portfolio.

Parmi les autres sources importantes exploitées, il faut mentionner l'analyse de tous les plans d'études cantonaux commanditée par la CDIP, des moyens d'enseignement relativement récents destinés à l'enseignement des langues étrangères, ainsi que des travaux de recherche théoriques et empiriques sur l'apprentissage des langues et l'évaluation des compétences en langues.

La contribution des résultats d'HarmoS pour la description des compétences s'inscrit dans le contexte d'un modèle de compétences déjà bien développé. Ce qu'il a surtout été possible de récolter grâce aux tests réalisés, c'est un grand nombre d'exemples de performances d'élèves dans les domaines de compétences de l'expression orale (*Prendre part à une conversation / S'exprimer oralement en continu*) et écrite, ainsi que des exercices de tests relatifs à la compréhension écrite, qui ont pu être classés selon les niveaux élaborés et qui permettent dorénavant d'illustrer concrètement les niveaux et leurs exigences (cf. le document mis en consultation relatif aux standards de base dans les disciplines «Langues étrangères»). Les descriptions de niveaux et donc aussi celles des standards se basent toujours sur les travaux réalisés antérieurement; les résultats des tests HarmoS ont conduit à des modifications et à des

⁶ Ces documents peuvent être consultés sur le site Internet du Conseil de l'Europe, à l'adresse suivante: www.coe.int/t/dq4/linguistic/CADRE_FR.asp.

ajouts dans les descriptions de compétences existantes, et constituent une source qui servira à concrétiser davantage les descriptions (cf. notamment le chap. 2.1 «Modèle de compétences» du présent rapport).

Dans ce contexte, il paraît utile de faire brièvement référence à des types d'échelles et de descriptions essentiellement différents (cf. Conseil de l'Europe, 2001: 35 sq.), qui peuvent être appliqués aux standards et à leur vérification:

- des échelles centrées sur l'utilisateur: elles décrivent des comportements typiques des candidats à n'importe quel niveau donné, en rendant compte, de manière principalement positive, de ce que quelqu'un est capable de faire.
- des échelles centrées sur l'examineur: elles expriment normalement les aspects qualitatifs de la performance attendue et ont ainsi pour fonction de guider la notation.

Les descriptions de niveaux utilisées dans le cadre du projet «HarmoS - Langues étrangères» comportent des éléments tirés de ces deux types d'échelles.

- des échelles centrées sur les exercices (échelles pour les concepteurs de tests): elles décrivent habituellement les tâches de communication qui doivent être assez directement mises en œuvre dans les exercices des tests.
- Spécifications centrées sur les tests et les exercices de tests (destinées aux concepteurs de tests): ces spécifications techniques règlent par exemple le volume du test et sa composition, des options concernant le format des exercices (quels formats ouverts et quels formats fermés?) et bien d'autres détails qui peuvent influencer le test et ses résultats de façon souhaitable (significative pour le *construct*⁷) ou non (pas significative pour le *construct*). Weir (2005b: 46) donne un aperçu des aspects qui doivent (devraient) en principe être réglés pour des exercices de tests relatifs à l'expression orale:

Caractéristiques de l'exercice: <ul style="list-style-type: none"> - But de l'exercice - Format de la réponse, attentes par rapport aux performances - Durée, longueur - Personnes impliquées - Critères linguistiques: <ul style="list-style-type: none"> - lexicaux - structurels - fonctionnels (par ex. types de dialogue) - Connaissance du monde présumée 	Interlocuteur(s): <ul style="list-style-type: none"> - Nombre - Vitesse d'élocution - Accent(s) - Degré de familiarisation - Sexe 	Conditions de réalisation: <ul style="list-style-type: none"> - Espaces utilisés (locaux) - Exécution concrète (préparation, etc.)
		Evaluation: <ul style="list-style-type: none"> - Examineur (cohérence, sévérité) - Critères d'évaluation

Ill. 3: Sources possibles de variance, significatives ou non pour le *construct*, dans les tests relatifs à l'expression orale (d'après Weir 2005b, 46)

Les enquêtes effectuées dans le cadre du projet «HarmoS - Langues étrangères» ont principalement porté sur les différents domaines de compétences qui relèvent des aptitudes langagières de communication (à savoir, concrètement: la compréhension de l'écrit, ainsi que l'expression orale et écrite). Toutefois, on a également retenu comme objectif pour les travaux de développement la nécessité de vérifier certains constats et, dans la mesure du possible, de dégager des enquêtes HarmoS des éléments pour les spécifications des tests et des exercices de tests, et ce pour deux raisons: d'une part, ces spécifications sont indissociablement liées à la validation du *construct* et à l'interprétation des résultats des tests, puisque les exercices des tests constituent une forme de transition déterminante entre la vérification et la description d'une compétence. En principe, la justification des interprétations tirées des résultats de tests constitue un défi primordial permanent, pour les personnes et les institutions chargées des tests de langues

⁷ Ndt: Le terme «construct» est d'usage courant dans la terminologie technique relative aux tests; il renvoie à la dimension latente que cherche à mesurer une épreuve grâce aux données observables que constituent les réponses des élèves.

HarmoS également. D'autre part, les informations ainsi récoltées sur les conditions, de même que sur le fonctionnement des tests et des exercices de tests s'avéreront utiles pour les enquêtes qui seraient réalisées ultérieurement ainsi que pour le futur monitoring de l'éducation. C'est pourquoi le présent rapport expose des aspects relatifs aux spécifications de tests et d'exercices de tests parallèlement à la présentation des enquêtes menées sur les différents domaines de compétences, par ex. la compréhension de l'écrit au chap. 3.3.1, et de manière plus étoffée au point 3.5.5.

2.2 Vérification du «construct» (méthodologie)

L'une des tâches principales du consortium Langues étrangères, dans le cadre d'HarmoS, a été de déterminer des domaines de compétences appropriés pour les groupes cibles, ainsi que de développer et de tester des procédures de vérification adaptées sur le plan didactique et du point de vue de la technique relative aux tests. Le point de départ conceptuel et pratique pour ces travaux a principalement été le *Cadre européen commun de référence pour les langues*, ainsi que les résultats de projets qui y ont fait suite, notamment *IEF / lingualevel*.

Un intérêt de recherche spécifique s'est dégagé, dans le contexte des enquêtes de validation, en rapport avec certains aspects à approfondir du modèle de compétences:

- Identifier des facteurs – ou des interactions entre des facteurs – qui influencent le degré de difficulté dans les tâches de communication (facteurs «task»);
- Reconnaître certains liens entre les dimensions de processus et d'action: dans quelle mesure les processus cognitifs qui interviennent dans l'exécution de certaines tâches (par ex. comportement de recherche, le fait d'interpréter, de tirer des conclusions) influent-ils sur l'issue d'une action réalisée dans des tâches de communication?;
- Identifier des effets liés à l'âge chez les élèves de sixième et de neuvième année.

Les travaux conceptuels et empiriques menés dans le cadre d'HarmoS visaient non seulement à opérationnaliser le point de départ qui sert de base de travail (*Cadre de référence, IEF / lingualevel*, notamment), mais aussi, autant que possible, à soumettre cette base à un examen critique, à la compléter et, le cas échéant, à la modifier.

Cependant, comme l'on pouvait s'y attendre étant donné le redimensionnement du projet et l'état de la recherche, ces objectifs ambitieux n'ont pu être que partiellement atteints.

L'opérationnalisation des descriptions de compétences du *Cadre de référence*, sous la forme d'exercices de tests, est un domaine dans lequel il reste passablement de travaux de développement à réaliser. De nombreux projets sont en cours actuellement, à l'échelle internationale, dans le domaine de l'opérationnalisation des descriptions de compétences par niveaux destinées à l'évaluation, surtout en rapport avec le manuel cité ci-dessus pour la classification des examens de langues en fonction du *Cadre de référence* (Conseil de l'Europe, 2003; Conseil de l'Europe, 2009). Les atouts des descriptions de compétences présentées dans le *Cadre de référence* (accessibilité aux non-professionnels, plausibilité, concision) sont parfois considérés comme étant justement des points faibles: les spécialistes en matière de tests notamment trouvent qu'il manque dans les descriptions une prise en compte systématique d'un ensemble défini de facteurs d'influence (Weir 2005a); ils déplorent le manque d'inventaires de textes, de critères textuels, d'éléments lexicaux et grammaticaux, de listes de conditions générales favorables ou défavorables, etc., dont ils pourraient tirer profit dans l'élaboration de tels exercices. Quoi qu'il en soit, de tels critères ne constituent pas des valeurs statistiques, et ils ne peuvent guère, en tant que tels, être attribués à un niveau précis. Il faut plutôt considérer que des interactions complexes se produisent entre, d'une part, les critères textuels et les paramètres afférents aux exercices et, d'autre part, les compétences des apprenants dans l'exécution de tâches de communication (cf. notamment Bygate et al. 2001).

De fait, il est question non seulement de comprendre le fonctionnement des tests, mais aussi de dégager, à partir de performances d'élèves dans des tests concrets, des indications permettant de conclure à la présence d'une compétence donnée qui, dans un certain cadre, peut aussi être

généralisable⁸. Par conséquent, il s'avère nécessaire, malgré l'évidente complexité de la relation élèves-exercices, de procéder par une approche fondée sur certains principes. En fin de compte, il s'agit, comme c'est d'ailleurs le cas pour toute évaluation, de garantir la validité du *construct* et, en principe, également de la démontrer.

Les personnes chargées d'élaborer des tests dans le domaine de l'instruction se réfèrent souvent, dans leur compréhension de la validité, à Messick⁹. D'après ce dernier, ce qui est fondamental pour la validité des tests, c'est d'éviter une sous-représentation du *construct* et une variance qui ne serait pas significative pour le *construct*, ainsi que d'assurer une qualité de tests adéquate, du point de vue de la technique des tests, et la pertinence des conclusions tirées à partir des résultats obtenus dans les tests (Messick 1996: 10; 12 sqq.). Il y a sous-représentation lorsqu'un test ne «couvre» pas le *construct*; c'est le cas par exemple quand les items ont été sélectionnés de façon peu diversifiée, ou qu'ils ne figurent pas en nombre suffisant dans le test. Des données lacunaires dans ce sens ne permettent guère de tirer des conclusions sur le *construct* en tant que tel. Une variance qui n'est pas significative pour le *construct* se produit par exemple lorsqu'un format de test nécessite la maîtrise d'une certaine capacité qui n'est pas répartie également dans la population et qui n'a rien à voir avec le *construct* à mesurer; des manipulations informatiques compliquées dans un test de langues entraîneraient probablement une variance indésirable. De plus, une variance qui n'est pas significative pour le *construct* peut aussi être causée par des formes d'évaluation et d'examen qui ne seraient pas suffisamment objectives, comme cela peut se produire avec des examens oraux (pour un développement plus complet à ce sujet, se référer à Lenz 2006: 210 sq.).

Nous avons cherché, pour les travaux qui ont été menés dans le cadre d'HarmoS, à satisfaire de diverses manières aux exigences posées par Messick. Nous reviendrons à chaque fois sur les détails de notre approche dans les sections qui traitent des travaux de développement réalisés dans les différents domaines de compétences retenus.

A l'heure actuelle, de nombreux examens qui se réclament du *Cadre de référence* testent les compétences linguistiques (ou plus précisément: le degré de *language proficiency* démontré dans l'utilisation de la langue) par le biais principalement de tâches de communication (*communicative tasks*) spécifiques aux aptitudes langagières (*skill-specific*); c'est le cas notamment des examens de certification de *Cambridge*, *DELTA* / *DALF* ou encore *Goethe*. Le fait de diviser la compétence d'utilisation de la langue, selon des aptitudes langagières de communication, en actes de compréhension orale, d'interaction orale, etc. a largement fait ses preuves dans la pratique. Cette systématique comporte toutefois un certain risque: celui que les actions plus englobantes, qui impliquent des activités langagières de différents domaines d'aptitudes, ne soient plus perçues comme une unité et ne soient plus testées comme telles. Dans la pratique, il arrive par exemple que certaines aptitudes se combinent dans ce sens lorsque le sujet réunit des informations à partir de sources orales et écrites pour ensuite les retransmettre sous forme écrite ou orale, ce qui, dans certains contextes, peut constituer une tâche tout à fait pertinente. Le consortium HarMoS Langues étrangères s'inscrit, lui aussi, avec ses tests dans la tradition d'une **évaluation de compétences langagières communicationnelles axée sur la tâche**¹⁰. Concrètement, les tests

⁸ C'est-à-dire des indications en rapport avec le *construct*: sur les domaines latents de savoir et de savoir-faire que l'on cherche à appréhender.

⁹ Les références fréquemment citées dans le domaine du *language testing* spécifiquement sont Bachman et Palmer, avec leur notion qualitative de ce qu'ils appellent la *test usefulness* (Bachman & Palmer 1996, 17 sqq.). Le concept de *test usefulness* y est compris comme un ensemble qui réunit et met en regard différents aspects qualitatifs de tests qui, dans d'autres conceptions, ont parfois été interprétés comme des antagonismes: fiabilité, validité du *construct*, authenticité, interactivité (les compétences visées sont-elles impliquées et activées?), *impact* et praticabilité. Ces deux derniers constituent des critères de qualité externes au texte, qui concernent les conséquences sur l'enseignement et l'apprentissage des langues (*impact*), et donc les limites, de la faisabilité pratique (dans ce modèle, le critère le plus élevé dans la hiérarchie).

¹⁰ Parmi les raisons qui nous ont poussés à opter pour une orientation sur la tâche, il y en a également d'ordre didactique: dans le domaine de la didactique des langues étrangères de manière générale, c'est en effet une approche très prometteuse, appelée à se développer, qui a entre-temps été relativement bien étudiée par la recherche et qui, dans les régions germanophones, a aussi trouvé, depuis quelques années, un ancrage toujours plus fort dans les contributions programmatiques et les travaux de recherche menés dans ce domaine (cf. Bygate, Skehan & Swain, 2001;

ont porté sur les compétences dans les domaines d'aptitudes que sont la compréhension de l'écrit ainsi que l'expression écrite et, dans une bien moindre mesure, l'expression orale. Cependant, des descriptions de niveaux sont aussi proposées, sur la base d'enquêtes antérieures, pour la compréhension de l'oral et, à titre d'essai, pour la médiation linguistique. Conformément au modèle des aptitudes, les compétences des élèves et donc les standards de base également ne sont pas décrits de manière *globale* pour l'allemand, l'anglais ou le français; les descriptions se présentent plutôt en rapport avec, par exemple, la compétence en compréhension écrite ou en rédaction dans les différentes langues.

Afin de garantir une meilleure couverture du *construct*¹¹ dans chacun des domaines d'aptitudes, on a en outre introduit, pour la conception des tests, des catégories linguistico-fonctionnelles, qui ont permis d'établir un classement des exercices de tests et de leurs différentes parties. On dénombre ainsi cinq *types de discours* (également appelés *fonctions rhétoriques*): informer / décrire, inviter / motiver à faire quelque chose, raconter / rapporter, exprimer son opinion / argumenter, maintenir la relation. Le but de ces catégories de classement était notamment d'éviter que les tests d'écriture ne contiennent que des (parties d')exercices où il est question de descriptions ou de récits; il s'agissait donc d'en inclure aussi qui nécessitent que l'élève exprime son avis et le justifie. Dans la compréhension de l'écrit, on a ajouté des distinctions fondamentales qui pourraient s'avérer significatives pour le *construct*, sur la base des différences établies pour les processus de lecture de PISA.

Les exercices de tests axés sur l'action qui sont utilisés dans le cadre du projet HarmoS sont en principe de type «holistique»; ils font intervenir un grand nombre de compétences, avec une complexité qui ne se laisse que difficilement appréhender. Et pourtant, il s'avère judicieux de procéder de la sorte, parce qu'on considère qu'une évaluation réductionniste, limitée à des compétences isolées (*discrete-point testing*), livrerait certes des résultats plus clairs à première vue, mais ne comporterait pas un potentiel de généralisation permettant, à partir des résultats «discrets», de dégager des assertions sur le comportement des personnes testées, notamment pour ce qui concerne les exigences de communication extrascolaires. Les exercices de tests qui présentent déjà des parallèles essentiels avec des tâches de communication telles que l'on peut en rencontrer en dehors de toute situation d'évaluation sont plus susceptibles de conduire à des extrapolations plausibles.

Lors de la conception des exercices, nous avons non seulement tenu compte des critères fonctionnels décrits ci-dessus, mais nous avons aussi pris en considération des aspects qui se rapportent aux compétences générales et aux compétences langagières communicationnelles citées plus haut. Les exigences liées aux compétences générales des élèves (comme la connaissance du monde présupposée ou la complexité cognitive d'un thème) ont été, dans la mesure du possible, autant contrôlées que celles qui relèvent du domaine des compétences langagières communicationnelles (par ex. le degré de différenciation du vocabulaire et la complexité des structures dans les textes de lecture). De plus, les ressources linguistiques ont joué un rôle important, comme critères, dans les grilles permettant d'évaluer la qualité des performances orales et écrites.

Le modèle de compétences qui a été pris comme point de départ pour les tests HarmoS, dans la lignée de la tradition du *Cadre européen commun de référence*, peut être présenté brièvement ainsi:

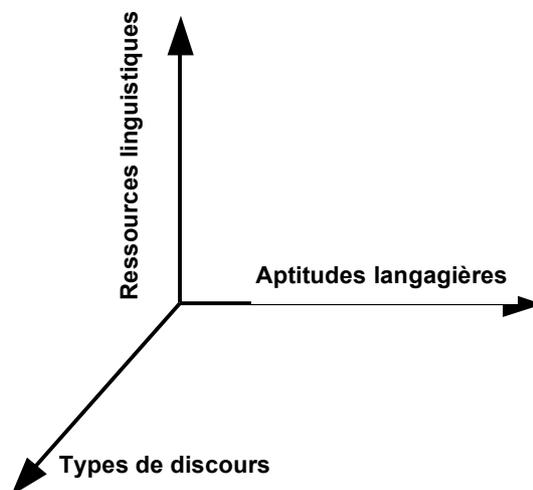
On cherche à tester la capacité du sujet à poser des actes de communication, dans le cadre d'activités langagières pouvant être classées dans les domaines d'aptitudes (*skills*) communicationnels suivants: s'exprimer oralement en continu, prendre part à une conversation, comprendre des textes et écrire. Les activités langagières suscitées par les exercices de tests requièrent la mobilisation de capacités dans le domaine des cinq types de discours suivants: informer / décrire, inviter / motiver à faire quelque chose, raconter / rapporter, exprimer son

Skehan & Foster, 2001; Bachman, 2002; Ellis, 2003 et 2005; Nunan, 2004; Robinson, 2005; Bausch et al., 2006; Müller-Hartmann & Schocker-v. Dittfurth 2005; Portmann-Tselikas, 2006; Long 2007).

¹¹ Par «couverture du *construct*» on entend une représentation suffisante (dans l'ensemble du test) des aspects (de compétences) que l'on cherche à étudier au travers du test.

opinion / argumenter, de même que maintenir la relation. Les exercices (et partant les critères d'évaluation) sont de plus caractérisés par diverses exigences, notamment dans le domaine des ressources linguistiques et communicationnelles (compétences linguistiques, sociolinguistiques et pragmatiques).

Ces aspects essentiels du modèle de compétences opérationnalisé dans le test peuvent être représentés graphiquement de la façon suivante:



III. 4: Les trois aspects centraux du modèle de compétences opérationnalisé

Il a fallu déployer des moyens considérables pour procéder à l'analyse statistique des données collectées – passablement volumineuses et de nature très diverse – ainsi que pour interpréter, sur cette base, les résultats des tests HarmoS. Une analyse et une interprétation qui n'ont, de plus, pas toujours pu être menées avec la profondeur et l'étendue souhaitées. En conséquence, il serait nécessaire de réaliser d'autres travaux d'analyse et d'interprétation, notamment en rapport avec la question, essentielle pour la compréhension des compétences, de savoir dans quelle mesure les différents domaines de compétences (les diverses aptitudes langagières de communication reprises dans HarmoS, et surtout la compréhension de l'écrit et l'expression écrite) ainsi que d'autres catégorisations (telles que celle selon les types de discours ou les processus de lecture) se profilent sur le plan statistique comme des dimensions empiriques distinctes. On peut par exemple se demander si une personne qui sait mieux lire que la moyenne sait aussi généralement mieux écrire que la moyenne – ou ces deux aptitudes sont-elles réparties différemment dans nos groupes cibles? En clair, forment-elles deux dimensions de compétence? La compétence lexicale se développe-t-elle parallèlement à la compétence grammaticale, ou certains maîtrisent-ils mieux la première et d'autres la seconde? Est-ce que certains élèves trouvent plus facile de lire des textes argumentatifs, et d'autres des textes informatifs? On ne dispose actuellement que de connaissances fragmentaires sur les points mentionnés ci-dessus comme des intérêts de recherche. Les interactions complexes qui surviennent entre les facteurs «personnes» et les facteurs «tâches» ne sont pas faciles à appréhender; il n'est pas aisé de donner une interprétation explicative des résultats empiriques. Par ailleurs, les conditions dans lesquelles s'est déroulée l'enquête posent aussi clairement des limites à un certain potentiel de connaissance. Dans ce contexte, il est particulièrement regrettable qu'aucune enquête qualitative n'ait pu – contrairement à ce qui avait été prévu en 2006 – être réalisée sur le fonctionnement des exercices. De fait, les élèves continuent à être une sorte de boîte noire; on ne dispose que de quelques rares données sociologiques susceptibles de constituer des facteurs «personnes»¹² qui pourraient être interprétés en conjonction avec des facteurs «tâches» et des caractéristiques de performances.

¹² Lenz (2006: 214-219) propose un aperçu portant sur un concept de validation plus complet.

3 Les enquêtes empiriques menées dans le cadre d'HarmoS

3.1 Vue d'ensemble

Dans un premier temps, le consortium Langues étrangères a pris part à l'enquête représentative principale qui a été organisée par la direction de projet HarmoS et réalisée au printemps 2007. Puis, il a également mené indépendamment des tests d'expression orale sur le terrain au mois de mai 2007. Dans les deux cas, seules la Suisse romande et la Suisse alémanique ont participé aux enquêtes; ces dernières n'ont pas été réalisées dans les régions italophones de Suisse. L'enquête principale comportait uniquement des tests écrits relatifs à la lecture et à l'écriture; on a également introduit, à titre d'essai, pour chaque langue, un test C – une sorte de test lacunaire – d'une durée de 20 minutes. Mais il restait trop peu de temps à disposition, après les schémas de planification infructueux de 2006, pour développer (et pour exécuter) des tests relatifs à la compréhension orale.

La deuxième langue nationale (allemand ou français) a été testée tant en sixième qu'en neuvième année; l'anglais, seulement en neuvième, étant donné que les élèves qui accomplissaient leur sixième année en 2007 n'avaient pas encore suivi de leçons d'anglais à l'école primaire.

Dans tous les domaines de compétences (y compr. le test C), une partie des exercices n'a été proposée qu'en sixième ou qu'en neuvième année, une autre dans les deux niveaux. Ce recoupement des exercices constituait une condition indispensable à l'élaboration d'une échelle (empirique) continue, qui permette de mettre en regard, en fonction de la langue cible, tous les exercices proposés et tous les élèves ayant participé. Du fait qu'on n'a pas cherché à tester des matières curriculaires et des contenus d'apprentissage (*achievement testing*), mais plutôt à évaluer les élèves en fonction de certaines qualifications (*proficiency testing*) – c'est-à-dire en fonction de ce qu'ils sont en mesure d'accomplir en rapport avec certains objectifs opérationnels – il a été plus facile de développer, ainsi que de mettre en œuvre et en échelle, des exercices applicables à plusieurs niveaux.

Un grand nombre d'exercices a été utilisé dans les trois langues cibles, moyennant quelques minimes adaptations qui s'avéraient nécessaires. Toutefois, il a été passablement difficile d'établir et de démontrer des comparaisons entre les résultats obtenus dans les trois langues cibles. Outre le recours à des méthodes statistiques (surtout pour la lecture), on a aussi opté pour certaines formes d'évaluation comparative, assurée par des experts (par ex. pour les performances relevant de la production). Cependant, il serait souhaitable de mener des travaux de recherche supplémentaires pour vérifier les résultats obtenus de la sorte.

A part l'élaboration d'une échelle interne à HarmoS, l'un des enjeux centraux a été d'établir (à nouveau) des liens entre les résultats obtenus et l'échelle de référence européenne, illustrée par des descripteurs. Cette démarche a démontré que se référer à cette échelle, pour chaque domaine de compétences et chaque langue, peut en fait constituer une étape supplémentaire, voire un autre moyen important, pour comparer entre eux les résultats des tests qui ont été obtenus dans les différentes langues.

Le fait de mener des tests en parallèle dans trois langues étrangères avec différents groupes dans plusieurs parties du pays rend le projet certes très laborieux, mais offre aussi, du point de vue scientifique et dans la perspective de la validation des tests, un correctif bienvenu qui permet d'éviter de tirer des conclusions hâtives à partir de résultats qui apparaissent, par exemple, dans les tests d'une seule langue.

3.2 Ampleur des enquêtes

Enquête représentative principale: avril-mai 2007

Les compétences en langues étrangères ont été testées dans la même mesure en Suisse romande et en Suisse alémanique. Le jour des tests, les élèves ont tous reçu quatre cahiers (2 cahiers de 30 min. et 2 cahiers de 20 min.), contenant des exercices portant sur la langue scolaire locale, sur la deuxième langue nationale et / ou sur l'anglais. La rotation des cahiers s'effectuait au niveau de chaque élève (à l'exception du test de compréhension orale relatif à la

langue scolaire locale, que toute la classe a fait en même temps, au début de la journée). Quelques rares restrictions ont été définies pour la répartition des exercices; par exemple, il ne fallait pas qu'un élève fasse des exercices sur la même aptitude langagière de communication (lire / test C, écrire) dans plus de deux cahiers, les trois langues possibles confondues. Ou encore: une langue étrangère devait toujours être testée pendant 30 minutes au moins, jamais uniquement avec un cahier de 20 minutes.

L'ensemble des tests a été réparti, à l'aide d'une structure matricielle, en fonction de tous les participants, si bien qu'à chaque fois, l'élève ne travaillait sur des cahiers de test que pendant 100 minutes d'affilée. De manière générale, il a été possible – grâce à une répartition (rotation) adéquate des cahiers et à un logiciel qui peut fonctionner avec des données très lacunaires (logiciel *Rasch* «ConQuest») – de collecter des informations qui, dans une configuration sans lacunes, auraient nécessité un temps de test individuel bien plus important. Les tableaux suivants proposent un aperçu des temps de test totaux (pour l'ensemble des tests virtuels), de même que du nombre de cahiers et d'exercices proposés¹³:

Durée totale pour les tests en 6^e année (deuxièmes langues nationales: allemand ou français)

	Français			Allemand		
	Durée totale	Nb. ex.	Durée totale	Nb. ex.	Anz. Aufg.	Anz. Hefte
Compréhension de l'écrit	140 min.	13 ex.	5 1/3 cahiers	140 min.	13 ex.	5 1/3 cahiers
Ecrire	120 min.	6 ex.	6 cahiers	120 min.	6 ex.	6 cahiers
Test C	20 min.	3 textes / 60 lacunes	2/3 cahier	20 min.	3 textes / 60 lacunes	2/3 cahier
Total	280 min.			280 min.		

Durée totale pour les tests en 9^e année (deuxièmes langues nationales: allemand ou français)

	Français			Allemand		
	Durée totale	Nb. ex.	Durée totale	Nb. ex.	Durée totale	Nb. ex.
Compréhension de l'écrit	130 min.	10 ex.	5 cahiers	130 min.	10 ex.	5 cahiers
Ecrire	130 min.	6 ex.	6 cahiers	130 min.	6 ex.	6 cahiers
Test C	20 min.	3 textes / 60 lacunes	1 cahier	20 min.	3 textes / 60 lacunes	1 cahier
Total	280 min.			280 min.		

Durée totale pour les tests en 9^e année (anglais)

	Anglais Suisse alémanique			Anglais Suisse romande		
	Durée totale	Nb. ex.	Durée totale	Nb. ex.	Durée totale	Nb. ex.
Compréhension de l'écrit	110 min.	9 ex.	4 1/3 cahiers	110 min.	9 ex.	4 1/3 cahiers
Ecrire	150 min.	7 ex.	7 cahiers	150 min.	7 ex.	7 cahiers
Test C	20 min.	3 textes / 60 lacunes	2/3 cahier	20 min.	3 textes / 60 lacunes	2/3 cahier
Total	280 min.			280 min.		

¹³ Pour la compréhension de l'écrit, ce qu'on entend ici par «exercice», c'est une combinaison de texte(s) et d'exercices (en moyenne 5 items environ) qui sont reliés par une même mise en situation. Pour ces exercices de lecture, on prévoit une durée de travail de 10 ou de 15 minutes. Dans le domaine de l'écrit, un exercice équivaut à rédiger un texte à partir d'une donnée. Le temps de travail s'élève à 20 minutes, et à 30 minutes pour un exercice proposé en neuvième année.

Les tableaux suivants donnent un aperçu du nombre d'élèves ayant pris part aux tests ainsi que de cahiers de test retournés.

Participation des élèves

	CH além.+CH rom. en tout	Suisse alémanique		Suisse romande	
		6 ^e année	9 ^e année	6 ^e année	9 ^e année
Nombre d'élèves qui ont retourné des cahiers dans les langues étrangères	5412	980	1512	1160	1760

Cahiers de test rendus

	6 ^e année		9 ^e année	
	Nb. de cahiers reçus	Cahiers reçus (%)	Nb. de cahiers reçus	Cahiers reçus (%)
Compréhension de l'écrit / test C - 2 ^e langue nationale (D ou F)	1830	90.4%	1311	86.9%
Compréhension de l'écrit / test C - anglais			1221	86.5 %
Ecrit – 2 ^e langue nationale	1122	75% ¹⁴	1101	81.8 %
Ecrit – anglais			1309	84.8%

Le taux enregistré dans les langues étrangères se situe un peu en dessous de ce qui est habituellement attendu dans des enquêtes de grande envergure de ce type. Quant au taux relevé dans les langues de scolarisation régionales testées simultanément (F et D), il est supérieur à 90%. Cette différence suscite la surprise. Nous ne disposons d'aucune information susceptible d'expliquer le taux plus faible constaté dans les langues étrangères (cf. toutefois la note de bas de page relative à l'écrit).

Tests d'expression orale menés sur le terrain en mai 2007

Le consortium Langues étrangères a mené de sa propre initiative des tests d'expression orale sur le terrain. Le but premier de ces enquêtes était de saisir un large spectre de performances d'élèves et de les documenter par vidéo, de façon à constituer une base solide pour débattre et illustrer des standards de base¹⁵.

La sélection des élèves n'est pas représentative, même si nous avons obtenu une représentation régulière des régions linguistiques et une distribution aussi large que possible au sein des cantons et des niveaux scolaires (sec. A-C); il a toutefois été difficile de trouver des volontaires dans les classes du niveau le plus bas (C). La distribution filles-garçons est équilibrée. En règle générale, nous avons filmé à chaque fois, dans une école donnée, trois élèves dans deux classes différentes. Ces élèves ont été choisis par l'enseignante ou l'enseignant parce qu'ils représentent les tiers inférieur, moyen et supérieur du spectre de performances de la classe (dans le domaine de l'expression orale). En tout, 174 élèves ont participé à l'enquête (F: 60; D: 54; E: 60).

¹⁴ Le nombre de travaux relatifs à l'écrit en 6^e année est plus faible uniquement parce que les résultats obtenus pour l'un des six exercices proposés ont été exclus de l'analyse. L'exercice en question n'a, contre toute attente, pas pu être interprété correctement par les élèves.

¹⁵ Seuls ont été filmés les élèves dont les parents avaient consenti à l'utilisation des enregistrements dans le cadre du débat qui a porté sur les standards.

Le tableau suivant donne un aperçu global des classes qui ont pris part au projet:

	Français		Allemand		Anglais	
	6 ^e année	9 ^e année	6 ^e année	6 ^e année	9 ^e année	6 ^e année
Elèves par niveau	30	30	30	24	30	30
Nombre d'écoles / classes	5/10	5/10	5/8	4/8	5/10	5/10
Types de classes	Primaire	Sec I, A: 2 Sec I, B: 5 Sec I, C: 3	Primaire	Sec I, A: 2 Sec I, B: 5 Sec I, C: 1	Sec I, A: 4 Sec I, B: 4 Sec I, C: 2	Sec I, A: 6 Sec I, B: 3 Sec I, C: 1

De ces tests sur l'expression orale ont résulté des enregistrements d'une durée brute (y compr. brève préparation avant les exercices) de plus de 50 heures. Les performances d'élèves saisies illustrent un spectre de niveaux allant de A1.1 à B1.2; certains élèves se situent même légèrement en dessus, alors même qu'aucun élève bilingue, au sens strict du terme, ne se trouvait parmi les participants.

Dans la planification initiale du projet, il était prévu, en plus de la participation à l'enquête représentative principale et des tests relatifs à l'expression orale menés sur le terrain, de réaliser des enquêtes complémentaires dans des classes spéciales ou des classes pilotes (par ex. celles bénéficiant d'un enseignement CLIL ou d'un début plus précoce), dans le but de réunir davantage de points de repère sur les compétences des élèves (pouvant être attendues à l'avenir). Toutefois, de concert avec la direction de projet HarMoS, nous avons décidé de ne pas poursuivre ce volet. Des enquêtes complémentaires seront envisagées à un stade ultérieur, lorsque le scénario 2010-2012 aura été mis en œuvre.

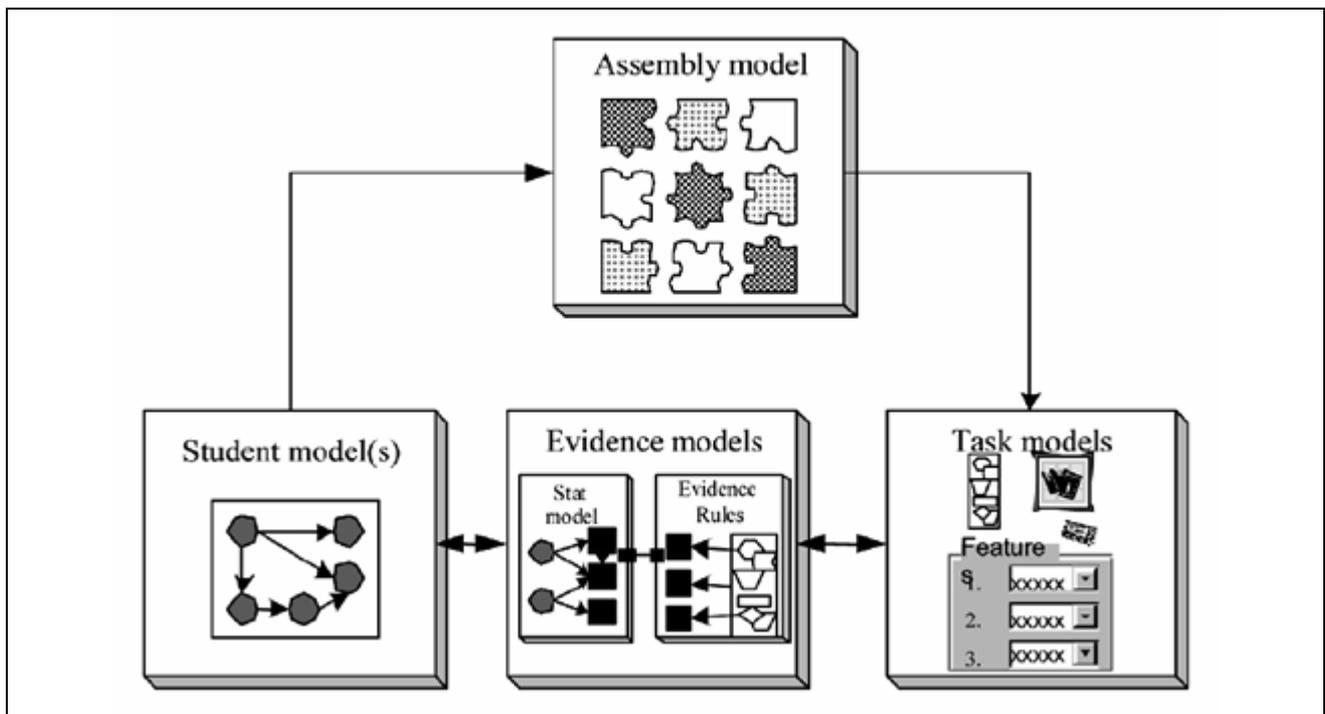
3.3 Instruments de test

3.3.1 Réflexions fondamentales sur l'évaluation des compétences

Les tests dans HarMoS ont pour fonction d'opérationnaliser un modèle de compétences, de façon à pouvoir saisir des performances, qui peuvent à leur tour se rapporter à ce modèle de compétences. Pour prouver la légitimité de cette démarche, il faut, dans le cadre d'une validation de tests, établir des liens, dans le cadre du modèle, entre les exercices, les performances, l'analyse des performances et les constatations qui s'en dégagent.

L'une des bases conceptuelles utilisées pour l'élaboration des tests dans le cadre du projet «HarMoS - Langues étrangères» a été le modèle *Evidence-Centred Design* (ECD – Mislevy, Steinberg, Almond & Russell 2002; Mislevy, Almond & Lukas 2004)¹⁶, un modèle d'enquête et de validation très courant, qui est aussi régulièrement employé dans le domaine de l'évaluation des compétences en langues (Mislevy et al. 2002; Bachman 2002).

¹⁶ Citons un autre modèle intéressant dans ce contexte, celui des *Four building blocks* de Mark Wilson (2004), qui ressemble fortement à celui de l'ECD.



Ill. 5: Schéma tiré du modèle *Evidence-Centred Design* (Mislevy et al. 2002: 481)

Le modèle ECD peut être interprété et utilisé ainsi: les variables significatives (savoir et / ou savoir-faire) sont spécifiées¹⁷, dans le *student model*, comme constituant des éléments ou des domaines de compétences. Le choix de ces éléments est souvent influencé par des réflexions externes (Que veut-on / doit-on savoir sur la population?) ou, comme dans le cas d'HarmoS avec le *Cadre de référence*, par une certaine compréhension des compétences. Les variables du *student model* sont des variables latentes, qui doivent être reliées à des variables observables:

*Operationally, SM [student model] variables summarize patterns of values of observable variables, along lines we build into the evidence model structures [...], as evoked by features we build into tasks*¹⁸.

Il est à noter que, dans le *student model*, on se concentre sur certaines compétences considérées comme centrales (c'est-à-dire significatives pour le *construct*). Il ne s'agit donc pas de l'ensemble des compétences dont dispose une personne et qu'elle peut, par exemple, mettre en œuvre pour exécuter des tâches (de communication) nécessitant l'utilisation de la langue, que ce soit en situation de test ou en dehors d'une telle situation. Néanmoins, il faut aussi prendre en compte, durant la phase de création des exercices et d'interprétation des performances, des difficultés survenant dans le domaine d'autres compétences (parmi lesquelles certaines ne sont pas significatives pour le *construct*)¹⁹.

Dans la mise en œuvre du *student model* sous la forme d'exercices, on cherche à créer des occasions, en nombre suffisant et de nature suffisamment variée, dans lesquelles il est nécessaire d'employer les compétences sur lesquelles se concentre le *construct*. En même temps, il faut aussi contrôler le recours à des compétences qui ne sont pas significatives pour le *construct* et

¹⁷ A ce stade, la *forme* de la description n'est pas encore définie. On peut trouver des indications sur des aptitudes abstraites (par ex. «compétence lexicale élevée») ou des généralisations sur une aptitude opérationnelle («Est aussi capable, la plupart du temps, de s'exprimer de manière appropriée en cas de...» – c'est ici que l'on classerait les descripteurs du *Cadre de référence*).

¹⁸ *Du point de vue opérationnel, les variables SM [student model] synthétisent des ensembles de valeurs de variables observables, selon des axes que nous intégrons dans les structures de l'«evidence model» [...], à l'instar des paramètres [features] que nous développons sous la forme de tâches [tasks]* (trad. libre d'une citation tirée de: Mislevy et al. 2002: 481).

¹⁹ C'est à partir de ces deux approches de la compétence que s'est fait jour, par exemple, le problème, essentiel pour l'interprétation des items de compréhension de l'écrit, de la divergence partielle entre le degré de difficulté tel qu'il est estimé (on s'oriente alors principalement selon des aspects linguistiques et communicationnels relatifs à la difficulté et à la performance) et le degré de difficulté empirique (qui se rapporte à des exigences / performances effectives dans le test, et par là même à des paramètres extérieurs au *construct*, qui ne sont pas significatifs pour celui-ci).

éviter des effets de distorsion, particulièrement lorsqu'il s'agit de difficultés qui peuvent être spécifiquement causées par la situation de test (par ex. avec des données d'exercice complexes, lorsqu'un appareil audio est de mauvaise qualité, ou que les contenus thématiques ne sont connus que de quelques élèves). Quoi qu'il en soit, l'une des difficultés majeures qui se présentent dans l'interprétation des résultats d'une enquête (ou de tests) réside dans la nécessité de déterminer la part relative des différentes ressources (compétences) dans les résultats obtenus, et ainsi de parvenir à des conclusions exploitables concernant le *construct* (sur les compétences) et concernant les compétences des personnes testées.

Outre le *student model*, on trouve un deuxième point de départ pour l'élaboration des tests, à savoir celui de la tâche (*task*), avec les modules que sont l'*assembly model* et le *task model*. Lorsqu'on opte pour une approche «à deux volets», comme l'a fait le consortium Langues étrangères en s'inspirant de Bachman & Palmer (1996: 12), il ne s'agit pas uniquement d'implémenter les éléments de compétences que l'on veut soumettre à examen (*student model*) en quantité suffisante dans les exercices conçus; il faut aussi le faire sous la forme d'exercices qui représentent les aspects particuliers de l'utilisation de la langue qui doivent être vérifiés (cf. à ce sujet, Lenz 2006: 214 sqq.). Il n'est normalement pas possible de procéder, dans des exercices de tests, à une mise en œuvre «naturaliste» de tâches reflétant des situations de la vie réelle. Les tests comportent certaines caractéristiques qui leur sont spécifiques et qui influent inévitablement sur les résultats. Dans les tests relatifs à la compréhension de l'oral et de l'écrit, il s'ajoute par exemple l'examen indirect de la compréhension au moyen de certains formats de réponse (par ex. questions à choix multiples, réponses brèves, évtl. résumés, en fonction des *task models* concrets qui sont choisis). Dans les exercices d'expression orale et écrite, on rencontre souvent un défi particulier dans la conception des tâches, à savoir la nécessité de formuler les données de façon à ce qu'une intention opérationnelle adéquate soit suscitée chez l'apprenant – un problème qui ne se présente souvent pas du tout dans les tâches qui surviennent dans des situations de vie réelle, mais qui peut influencer de façon déterminante la performance réalisée dans un test. Une «compétence» ne peut en principe être observée directement ni dans le cadre de tests, ni dans celui de situations de vie réelle; toutes les conclusions tirées sont donc le produit d'inférences. Or, pour permettre des inférences, il faut nécessairement que des éléments essentiels puissent d'abord être observés. Cela n'est possible que si les exercices mobilisent des aspects centraux de la / des compétence(s) visée(s) et ne produisent parallèlement pas d'effets perturbateurs assez forts pour détourner ou déformer le regard qui est porté sur la compétence essentielle (mais cachée) – comme ce serait le cas, par exemple, dans un test de compréhension orale qui vérifierait la compréhension de la personne en lui demandant de réaliser une performance écrite complexe.

Les *evidence models* ont pour fonction de saisir les paramètres importants qui se dégagent de la combinaison des exercices et des performances correspondantes, réalisées par les apprenants, ainsi que de les transférer en propositions concernant les compétences ciblées dans le *construct*, en l'occurrence dans le *student model*. A l'inverse, les *evidence models* fournissent également des points de repère essentiels permettant de savoir à quoi ressemblent des performances qui correspondent à certaines orientations du *construct* dans certains contextes – en faisant apparaître des résultats de tests et des caractéristiques de performances d'élèves auxquels on peut s'attendre. Cette double relation d'interprétation, qui s'établit au travers des *evidence models*, est une cheville importante du modèle ECD et constitue un défi permanent, surtout si l'on veut tester des compétences en langues sans devoir renoncer, d'une part, à des tâches de communication globales et, d'autre part, à des propositions portant sur un sujet aussi complexe que la compétence d'utilisation d'une langue²⁰.

²⁰ Dans un article de pointe sur le sujet, les auteurs Alderson & Banerjee dressent, après une vingtaine d'années consacrées à l'élaboration de tests de communication, un bilan plutôt sceptique:

Les stratégies et probablement les traits [c'est-à-dire les dispositions cognitives individuelles] peuvent varier selon les personnes et les tâches, même lorsque ces personnes obtiennent des résultats identiques. Un même résultat de test peut représenter différentes aptitudes, ou diverses combinaisons d'aptitudes, ou encore différentes interactions survenant entre des traits et des contextes, et il est actuellement impossible de dire précisément ce qu'un résultat signifie exactement. C'est ce qu'on pourrait appeler le «trou noir» des tests de langues (trad. libre tirée de: Alderson & Banerjee 2002, 100 sq.). Citation originale:

Mislevy et al. expliquent ainsi le rapport qui s'établit entre les compétences, les propriétés des exercices et les variables observables, donc entre les différents modules du modèle ECD:

Distinct SM variables are used to maintain belief about distinct knowledge / skills, and a claim is associated with particular patterns across them as they are called upon in tasks that stress competences in different ways. Thus, students' proficiency in a domain can be described in terms of which skills they possess (via SM variables), tasks can be described in terms of which skills they require (via TM [task model] variables) and the outcomes expected from any given match-up can be described in terms of values of observable variables²¹.

Comme on peut le constater en étudiant le modèle ECD, l'élaboration de tests et le développement ainsi que la validation d'un modèle de compétences constituent les éléments interdépendants d'un processus récursif; la validation du modèle et de l'instrument de recherche n'est en principe jamais achevée – et pourtant elle est indispensable, surtout pour les nouveaux outils développés. Dans ce contexte, on peut conclure – dans le sens de Messick – qu'il faut prendre des mesures de validation qui entretiennent un rapport approprié avec les conséquences (réelles) qui résultent des tests pour les personnes concernées.

3.3.2 *Elaboration des tests dans HarmoS*

Si le modèle ECD constituait la base méthodologique pour l'élaboration des tests, c'est le modèle de compétences du *Cadre de référence* qui a en principe été pris pour point de départ, au niveau du contenu, pour le développement des tests HarmoS relatifs à la compréhension de l'écrit et à l'écrit (enquête représentative), ainsi qu'à l'expression orale en interaction et en production (tests menés sur le terrain). Le modèle de compétences du *Cadre de référence* a été concrétisé dans différents documents, dans le *Cadre européen commun de référence* lui-même et dans des documents qui ont ensuite été réalisés sur cette base, parmi lesquels on trouve:

- le *Portfolio européen des langues* et des travaux réalisés sur le portfolio;
- les résultats du projet alémanique *IEF* (produit commercial *lingualevel*);
- des documents basés sur le *Cadre de référence*, qui le complètent et qui ont été réalisés dans des projets du Conseil de l'Europe (manuel pour la classification des examens de langues; la grille qu'on appelle *Dutch Grid* et qui permet de décrire des caractéristiques d'exercices significatives; des documents de l'*Association of Language Testers in Europe*).

Pour compléter cette base, nous nous sommes également référés à des résultats de recherche théoriques et empiriques portant sur l'apprentissage et l'évaluation des langues, dont notamment des travaux réalisés en rapport avec PISA (*reading literacy*).

Outre la décision de procéder à des tests différenciés par aptitudes langagières de communication (compréhension de l'écrit, etc.), et donc de traiter ces dernières comme des domaines de compétences distincts, il a aussi été question, assez rapidement, de veiller spécialement à couvrir suffisamment le *construct* dans le domaine fonctionnel. Dans ce but, on a procédé, en s'appuyant sur diverses classifications, différenciant légèrement les unes des autres, à une répartition fonctionnelle des exercices selon leurs types de discours dominants (informer / décrire, etc. – cf. à ce sujet le chap. 2.2, plus haut).

Un troisième domaine de catégories a été défini avec des composantes constitutives du *construct*; il comprend les différentes ressources du *Cadre de référence* qui permettent la mise en œuvre de

Strategies, and presumably traits, can vary across persons and tasks, even when the same scores are achieved. The same test score may represent different abilities, or different combinations of abilities, or different interactions between traits and contexts, and it is currently impossible to say exactly what a score might mean. This we might term The Black Hole of language testing.

²¹ *Des variables SM distinctes sont utilisées pour maintenir une perception de connaissances / capacités distinctes, et on associe des spécifications à certains schémas [patterns] récurrents dans ces variables, lesquelles interviennent dans des tâches qui mettent en évidence des compétences de différentes manières. Ainsi, le degré de maîtrise des étudiants dans un domaine peut se décrire à partir des capacités dont ils font preuve (et ce, par le biais des variables SM); les tâches peuvent être décrites sur la base des capacités qu'elles exigent (grâce aux variables TM [task model]), et les résultats que l'on peut attendre d'une combinaison donnée peuvent se comprendre en termes de valeurs de variables observables (trad. libre d'une citation tirée de: Mislevy et al. 2002: 484).*

compétences, c'est-à-dire les «compétences générales» et les «compétences communicatives langagières» (cf. chap. 2.1, ci-dessus).

D'autres approches de systématisation ont pu s'y ajouter, selon le domaine de compétences, pour l'élaboration des exercices, ce qui devait permettre de couvrir le *construct* avec une plus grande fiabilité (cf. notamment, à ce sujet, le classement des exercices en fonction de processus de lecture dans les tests de compréhension écrite).

Les différents éléments qui ont été retenus pour garantir une meilleure couverture du *construct* par les exercices ont en principe le potentiel d'apparaître, dans l'analyse des résultats empiriques, comme des catégories distinctes au niveau empirique, sinon on ne devrait pas en faire état séparément. Il doit donc être possible que, en fonction de ces catégories, ce soient à chaque fois d'autres élèves qui obtiennent de bons ou de moins bons résultats. Quant à savoir si la preuve empirique de la dimensionnalité selon ces catégories est apportée ou non dans un cas concret, cela dépend toutefois aussi passablement de l'ampleur des tests réalisés et de la qualité des données collectées.

Des *construct maps* ont été conçues pour servir de base à l'élaboration concrète des exercices de tests, mais aussi éventuellement à la définition des grilles d'évaluation appliquées dans les différents domaines de compétences (aptitudes langagières de communication). Ces «maps» ont permis de concrétiser les catégories mentionnées en fonction des niveaux fixés. Nous avons pris, comme principales références pour ces paramètres, les descriptions de niveaux étalonnées du *Cadre de référence* et du projet *IEF / lingualevel*, lesquelles ont été évaluées en fonction des différents éléments en jeu. Les chapitres suivants, qui portent sur les différents domaines de compétences, traitent cette question plus concrètement.

Nous avons également pris des décisions concernant les formes et les formats des exercices:

- Seules des tâches de communication (*communication tasks*) sont proposées comme instruments de test permettant d'évaluer les différents domaines de compétences.

De plus, on a introduit pour chaque langue testée un bloc de 20 minutes avec un test C, qui ne se rapporte pas d'emblée à un domaine de compétences précis parmi ceux qui font l'objet de tests. Le test C est une forme de test lacunaire, qui est proposé pour permettre une évaluation intégrative de la compétence langagière. On dit du test C qu'il est idéal pour obtenir, avec un degré d'efficacité et de fiabilité très élevé, une *vue d'ensemble* des compétences d'un individu dans une langue donnée, ou encore pour dresser un bilan des capacités déployées dans le domaine de l'écrit, et particulièrement l'aspect rédactionnel. Jusqu'à présent, le test C a surtout été étudié avec des adultes possédant des capacités en littérature bien développées, tels que des candidats se préparant à des études niveau haute école²². Le recours, dans le cadre d'HarmoS, à un test C (légèrement réduit) avait initialement pour but de sonder son potentiel (diagnostique) en rapport avec des apprenants plus jeunes. Cependant, il a fallu repousser provisoirement les analyses menées dans ce but.

- Les tâches de communication font l'objet d'une courte mise en situation à l'intention des élèves; la donnée établit un lien avec des situations plausibles en dehors du test, dans lesquelles cette tâche ou une tâche semblable peut se présenter. Cette mise en situation vise à susciter une attitude qui motive l'élève à exécuter la tâche proposée et l'aide à se concentrer sur ce qui est attendu de lui.
- La langue utilisée pour les explications, les données et les questions est en principe la langue de scolarisation de l'élève; ce choix s'explique par différentes raisons: premièrement, les exercices doivent être compris, dans une situation de test où l'élève ne reçoit aucune aide extérieure; deuxièmement, la compétence en lecture ne doit être testée que là où il est expressément question de la tester, et là encore, il s'agit de tester la compétence en lecture de manière ciblée, en rapport avec les textes proposés, et non avec les formulations contenues dans les exercices (par ex. les réponses à choix multiples); troisièmement, il s'agit d'éviter que

²² Le site Internet www.c-test.de donne un bon aperçu de la théorie et de la pratique liées au test C. La liste des références bibliographiques propose de nombreuses publications importantes traitant de ce thème, notamment celles de R. Grotjahn.

des élèves ne répondent juste à un exercice que grâce à des mots contenus dans la donnée (qu'ils n'auraient peut-être pas compris sinon).

- Pour l'expression orale et écrite, nous avons sélectionné des formats d'exercices très ouverts, c'est-à-dire pas des textes lacunaires ou des procédés comme les *prompts*, qui sont contenus dans la formulation, et auxquels on ne peut répondre que par un spectre de possibilités très limité.
- Dans les tests de compréhension écrite, on trouve toute une série de formats d'exercice, notamment dans l'idée de neutraliser tout effet de distorsion éventuel que pourrait avoir un certain format d'exercice, ou éventuellement de pouvoir l'identifier grâce à l'analyse des résultats. De manière générale, on a toutefois cherché à éviter des corrections trop importantes, donc à ne pas devoir procéder à des évaluations proprement dites de réponses ouvertes relativement longues.

Le nombre d'exercices (et d'items) possibles a été déterminé concrètement par le temps de test total mis à la disposition des langues étrangères. En procédant à des estimations par niveaux et par exercices, nous nous sommes assurés à l'avance qu'il y aurait des exercices appropriés pour l'ensemble du spectre des performances que réaliseraient les élèves participant aux tests. Afin de permettre la création d'une échelle de compétences continue, une partie des exercices de chaque domaine d'aptitudes a été proposée tant en sixième qu'en neuvième année. Les recoupements ont été plus importants dans les exercices d'expression orale et écrite, plus «larges» en niveau, que pour la compréhension de l'écrit.

Les exercices ont été développés en plusieurs étapes. Des sous-groupes travaillant par domaine ont fait de premières propositions, qui ont ensuite été présentées pour approbation à tout le consortium. Après une phase de remaniement, une sélection a été effectuée parmi les exercices et ceux qui ont été retenus ont alors été testés en classe dans des conditions temporelles très contraignantes. L'ampleur des tests préliminaires n'a pas été la même pour tous les domaines:

- Compréhension de l'écrit: CH além.: 7 enseignantes et enseignants - 8 classes; CH rom.: 8 enseignantes et enseignants - 9 classes;
- Ecrire: CH além.: 16 enseignantes et enseignants - 18 classes; CH rom.: consultation de 3 didacticiens spécialisés;
- Parler: différents élèves d'une école primaire et d'une école secondaire de Suisse alémanique; puis les exercices ont encore été testés, avec la participation de 6 élèves alémaniques et de 6 élèves romands, dans le cadre de la formation qui a été dispensée aux interlocuteurs.

A la suite de ces tests préliminaires, nous avons procédé à des analyses approfondies et à des remaniements dans les exercices.

La traduction des exercices a été assurée en partie par le consortium et en partie par la CDIP. Dans les exercices relatifs à l'expression orale et écrite, il s'agissait presque exclusivement de traductions d'allemand en français; aucun texte de départ n'a été fourni, qui aurait nécessité une traduction dans les trois langues cibles. Pour la compréhension de l'écrit, il a fallu, outre la traduction en français des instructions, des mises en situation et des données, traduire également une partie des textes de lecture. Les exercices ont toujours été construits sur la base de textes réels, dont l'original était soit en français, soit en allemand, soit en anglais. Il a aussi fallu entreprendre des traductions en fonction des autres langues cibles pour lesquelles ces textes étaient utilisés. L'enjeu de ces traductions a été de reproduire des textes ayant des propriétés comparables.

Suivit un processus laborieux pour les exercices sélectionnés pour l'enquête représentative, processus qui se déroula en collaboration avec des graphistes et l'éditeur. Puis il y a encore eu quelques difficultés de mise en route, qui ont pris passablement de temps, mais le résultat est tout à fait satisfaisant.

L'enquête représentative s'est déroulée sous la responsabilité de la CDIP. Les langues étrangères ont fait partie de l'enquête en Suisse alémanique et en Suisse romande, mais pas au Tessin. La réalisation concrète des tests n'a pas été suivie sur place; elle a eu lieu sous la responsabilité des écoles qui ont pris part à l'enquête. Il est difficile d'évaluer les conséquences qui en découlent; on

ne dispose par exemple que de quelques éléments qui indiqueraient que le temps imparti aurait été utilisé pour préparer concrètement les élèves aux exercices des tests.

Quant aux tests d'expression orale menés sur le terrain, ils ont été réalisés sur place avec la participation, à chaque fois, d'une équipe de trois personnes – deux locuteurs dont la langue maternelle était la langue testée, plus une personne du consortium, qui assumait la responsabilité organisationnelle et disciplinaire du projet. Les contributions des élèves ont toutes été filmées du début à la fin.

Les résultats de l'enquête ont été corrigés, évalués, puis codifiés par des personnes externes qui ont contribué à cette phase du projet. C'est le groupe «Méthodologie» qui a procédé à l'analyse statistique des résultats obtenus dans les tests représentatifs. Les résultats des tests d'expression orale menés sur le terrain ont été, autant que cela était possible, évalués du point de vue statistique en interne, au sein du consortium. L'analyse des données, notamment celles de la partie «Ecrire», s'est avérée si laborieuse qu'il a fallu un mandat supplémentaire pour pouvoir analyser de premiers résultats bruts de manière approfondie. De même, en ce qui concerne les autres parties, il n'a pas encore été possible de réaliser, dans l'étendue qui est souhaitée, les analyses d'approfondissement qui permettent d'explicitier les principaux résultats empiriques. Et pourtant, les résultats que l'on obtiendra dans ce domaine sont justement déterminants pour tirer des conclusions sur un domaine clé de la validation des tests, à savoir la relation qui s'établit entre: modèle de compétences / *construct* et test.

Les chapitres suivants développent divers aspects des enquêtes qui ont été réalisées dans les différents domaines de compétences. Ces développements suivent, pour l'essentiel, les points qui ont été traités dans cette introduction: compréhension des compétences / *construct*; élaboration des exercices, collecte des données; correction / évaluation; analyses et résultats. Chaque section présente enfin brièvement les perspectives qui s'ouvrent en vue des travaux qui doivent encore être menés dans le domaine²³.

3.4 Compréhension de l'oral

Aucune enquête n'a pu être réalisée sur la compréhension de l'oral dans le cadre du projet HarmoS²⁴. Or c'est un domaine central de la compétence en langue étrangère, qui joue un rôle clé notamment dans l'apprentissage précoce d'une langue. Dans l'utilisation de la langue, la compréhension de l'oral ne joue pas seulement un rôle important comme aptitude distincte (et comme compréhension audio-*visuelle*); elle intervient également dans l'interaction orale et la médiation linguistique. Par conséquent, la compréhension de l'oral a pleinement sa place dans le modèle de compétences «Langues étrangères», et c'est pour cette raison que des standards de base ont aussi été proposés pour ce domaine de compétences.

En vue de situer, ne serait-ce que grossièrement, la compréhension de l'oral, les sections suivantes présentent très brièvement et sous une forme synthétique, des positions choisies dans la littérature spécialisée, qui se rapportent à des parallèles et à des différences entre l'écoute et la lecture, de même qu'entre la compréhension orale dans la langue première et celle dans une langue étrangère. Ensuite, vu que les propositions pour les standards de base ont été illustrées par des exercices tirés du projet *IEF / lingualevel*, une section présente plus en détail les tâches de

²³ Bien que les différentes étapes de développement et de validation décrites puissent être rapportées au cadre conceptuel du modèle *Evidence-centred Design* (cf. 3.3.1), il ne sera plus fait explicitement mention de ce modèle dans les sections qui suivent.

²⁴ S'il a été possible d'observer subsidiairement l'aspect de la compréhension orale dans l'interaction directe qui s'est produite dans les conversations des tests d'expression orale menés sur le terrain, on a toutefois renoncé, pour des raisons de temps mais aussi pour des considérations méthodologiques, à procéder à une analyse de cet aspect.

communication relatives à la compréhension de l'oral qui sont proposées dans le cadre de ce projet en particulier²⁵.

Dans la littérature spécialisée, on considère qu'il existe des parallèles assez prononcés entre les processus intervenant dans la compréhension de l'oral et ceux qui se produisent dans la compréhension de l'écrit (cf. chap. 3.5). Cependant, il est admis que les «processus en temps réel» qui permettent de traiter des informations linguistiques revêtent une signification bien plus grande dans la compréhension orale que dans la lecture (cf. notamment Weir 2005, Rost 2002, Buck 2001, Solmecke 2000; cf. aussi Nold / Rossa 2006). Dans la compréhension orale, l'information à traiter est fugace et ne s'accompagne pas des mêmes possibilités de contrôle que dans la lecture. Car en lisant, on peut reconsidérer des passages difficiles plusieurs fois, les relire attentivement ou encore les sauter. Dans la compréhension orale, cela n'est pas possible, et des problèmes de compréhension se présentent rapidement, par exemple si quelqu'un parle vite ou peu distinctement, ou encore que des connaissances contextuelles font défaut. – Dans la compréhension orale en langue étrangère, les «processus en temps réel» nécessaires pour comprendre la langue sont encore plus susceptibles de connaître des perturbations que dans la langue première. Cela s'explique par le fait que les ressources langagières dans la langue étrangère sont souvent encore incomplètes, et moins rapidement mobilisables, que dans la langue première. A cela s'ajoute le fait que les stratégies d'adaptation de l'écoute dans une langue étrangère aux exigences situationnelles, aux textes oraux et aux données des exercices ne se transfèrent pas sans difficulté depuis la langue première; ces stratégies doivent souvent d'abord faire l'objet d'une prise de conscience dans la langue étrangère, et même être consciemment développées.

Dans le projet *IEF / lingualevel*, la compréhension de l'oral est conçue, en rapport direct avec le *Cadre européen de référence*, comme une activité communicationnelle de réception, qui peut être liée à différentes intentions d'écoute. Parmi ces buts, on trouve (Conseil de l'Europe, 2001, p. 54): comprendre «l'information globale» (saisir le sens global), comprendre «une information particulière» (obtenir une information très spécifique), comprendre «l'information détaillée» (comprendre ce qui est dit dans tous les détails), pouvoir tirer des conclusions, etc. Nous avons cherché à mettre en œuvre de tels buts dans les exercices conçus, grâce à des données appropriées, à des mises en situation adéquates, ainsi qu'à des textes et des formats adaptés. Toutefois, nous n'avons pas systématiquement fait varier les intentions d'écoute.

Les données des exercices sont largement reprises des descriptions de compétences pour l'écoute proposées dans le projet *IEF / lingualevel*; elles ont été formulées de manière à refléter autant que possible des contextes oraux du quotidien. Par des mises en situation adéquates, on cherche à produire des attitudes d'écoute aussi réelles que possible, qui peuvent ou pourraient s'avérer significatives pour les élèves. Voici des exemples de mises en situation: «Tu as enregistré une émission radio avec une jeune sportive»; «Tu veux savoir quelle est la météo dans certaines régions de Suisse allemande. Tu as déjà appelé une fois le service météorologique par téléphone et tu as pris des notes. Maintenant, tu rappelles et tu vérifies tes notes»; «Tu as été invité à une petite fête chez Julia, une amie de Lucerne. Tu entends quelques échanges courts; «Tu entends par hasard dans un magasin un jeune Suisse allemand qui commande un jeu d'ordinateur».

Lorsque c'était possible, nous avons toujours conçu les exercices d'écoute à partir de textes oraux authentiques, qui ont été trouvés tels quels ou qui ont ensuite été «reconstitués» en fonction du

²⁵ Le projet IEF (*Instruments pour l'évaluation des compétences en langues étrangères I* «Instrumente für die Evaluation von Fremdsprachenkompetenzen») a été lancé et soutenu financièrement par les conférences régionales alémaniques de la CDIP, à savoir: NW EDK, EDK-Ost et BKZ (coordination). Le projet a été mené par Peter Lenz et Thomas Studer, au *Centre d'Enseignement et de Recherche en Langues Etrangères* de l'Université de Fribourg (CH). Il s'est déroulé entre 2002 et 2005. Dans le cadre de ce projet, des instruments d'auto-évaluation et d'hétéro-évaluation ont été développés et testés, en collaboration avec de nombreux enseignants et enseignantes, ainsi qu'avec la participation de leurs classes, et ce pour l'enseignement du français et de l'anglais à partir de la 5^e et jusqu'à la 9^e année. Ces instruments peuvent être utilisés à des fins d'évaluation variées. Les instruments ainsi conçus comprennent des descriptions de compétences et des exercices de tests pour les cinq aptitudes que sont *écouter*, *lire*, *expression orale interactive*, *s'exprimer oralement en continu* et *écrire*, ainsi que des grilles d'évaluation et des performances de référence commentées pour les aptitudes de production. Ces instruments ont été classés, au moyen de méthodes scientifiques, en fonction des niveaux de référence du *Cadre européen commun de référence pour les langues* du Conseil de l'Europe. Ils ont été publiés sous le nom de *lingualevel* par les éditions scolaires bernoises *Schulverlag* (cf. www.lingualevel.ch). Lenz / Studer 2005 et Studer / Lenz / Mettler 2004 proposent des présentations plus générales du projet.

modèle utilisé. Tous les textes ont été prononcés par des locuteurs natifs, qui venaient de différentes régions dans lesquelles on parle la langue cible. On constate des variations assez importantes pour l'anglais notamment (Angleterre, Etats-Unis, Inde, Irlande). En français, c'est la diversité romande qui prédomine.

Les textes se rapportent premièrement à des domaines privés, scolaires et publics d'utilisation de la langue. Sont notamment représentés: des conversations du quotidien (interactions face à face, conversations téléphoniques et discussions entre plusieurs personnes), des émissions de radio, des entretiens et des annonces. Selon le domaine d'utilisation de la langue, l'input linguistique a tendance à être plutôt de type texte conceptuel oral ou texte conceptuel écrit. Les textes conceptuels oraux (tels que des conversations du quotidien) s'accompagnent, dans une assez grande mesure, de phénomènes spécifiques au langage parlé, comme des assimilations phonétiques, des ellipses, des phénomènes d'hésitation et des signaux de réception. Les textes conceptuels écrits (par ex. les nouvelles à la radio ou des annonces) sont orientés plus ou moins fortement selon des textes écrits et sont souvent caractérisés par des signaux de planification plus précise (notamment structure textuelle plus claire, renvois à la langue, souvent aussi syntaxe plus complexe). Dans l'ensemble, un large spectre de fonctions linguistiques est réalisé dans les textes oraux; on trouve des textes descriptifs et informatifs, tout comme des textes instructifs, narratifs et argumentatifs, sans oublier des textes de prise de contact et des échanges pour entretenir la relation.

Les exercices d'*IEF / lingualevel* comprennent les formats de réponse suivants: juste-faux, réponse brève (en partie ouverte, en partie orientée par des mots-clés ou des phrases) ainsi que questions à choix multiples. Le choix de ces formats constitue aussi en partie un compromis entre des exigences théoriques afférentes aux tests et le naturel de l'approche adoptée dans le projet *IEF / lingualevel*. Par exemple, on pourrait relever qu'avec le format juste-faux, la probabilité de deviner juste est relativement élevée. D'un autre côté, l'une des consignes *IEF / lingualevel* les plus courantes dans les items juste ou faux, à savoir le fait de vérifier des attentes ou des hypothèses (qui repose en situation réelle sur des connaissances préalables), reflète une attitude d'écoute réelle et tout à fait courante. C'est la raison pour laquelle le format juste-faux a également été utilisé.

Dans le but de limiter le risque d'une surcharge (cognitive) des élèves occasionnée par les exercices de compréhension orale, les élèves ont pu écouter les textes oraux deux fois (surtout lorsque ce serait aussi possible en dehors de la situation de test, par ex. avec des émissions enregistrées ou des annonces), ou alors des pauses ont été intégrées dans les textes oraux (notamment quand il fallait écrire une réponse courte). De cette manière, les élèves avaient le temps de répondre à l'exercice et de se préparer pour le texte oral suivant.

Toutes les consignes sont formulées en français (donc dans la langue de scolarisation de Suisse romande, et non dans la langue cible). Il s'agit premièrement de garantir la compréhension de l'exercice; deuxièmement, d'éviter que ce soit la compréhension de l'écrit plutôt que celle de l'oral qui soit testée, et troisièmement, d'empêcher que des réponses ne soient données que parce que des mots contenus dans la formulation de la consigne seraient reconnus dans le texte oral. De manière générale, le recours à la langue scolaire locale dans les exercices des tests constitue une mesure cohérente permettant d'accroître la validité des exercices. Il n'est pas du tout question d'opposer un mouvement inverse aux efforts, tout à fait positifs, déployés par de nombreux enseignants et enseignantes ou didacticiens et didacticiennes pour faire de la langue cible la langue de communication durant les cours.

Les exercices *IEF / lingualevel* relatifs à la compréhension de l'oral ont été développés entre 2004 et 2005. Ils ont été validés tant du point de vue qualitatif que par des méthodes statistiques. La création des exercices et leur validation ont, pour l'essentiel, comporté les étapes suivantes:

- étude de la littérature spécialisée et analyse d'exercices déjà existants (dont des exercices tirés d'ouvrages d'enseignement, de recueils d'exercices et d'examens de certificats);

- élaboration de premiers exercices sur la base d'un ensemble de descripteurs, qui ont été conçus durant la première phase du projet *IEF*²⁶; production de textes oraux; définition de dispositions d'exécution et de clés de résolution;
- vérification qualitative (test) des exercices dans des classes; remaniement des premières propositions d'exercices; développement du corpus d'exercices et des clés de résolution; finalisation des dispositions d'exécution;
- exécution des tests de compréhension orale dans un total d'environ 85 classes, avec des élèves d'âge et de niveau différents. Nombreux sont les élèves qui ont non seulement répondu aux exercices, mais qui ont en plus rempli un questionnaire d'auto-évaluation, composé de descriptions de compétences IEF, lesquelles avaient déjà auparavant été classées en fonction des niveaux de référence.
- calibrage des données relatives aux tests et à l'auto-évaluation avec *Winsteps* et *Facets*. Elimination des items problématiques. Sur la base des auto-évaluations remplies, classement des exercices selon les niveaux de référence.

Ces travaux de développement ont débouché sur 75 exercices de tests validés de manière empirique (460 items validés), dont environ la moitié pour le français et l'anglais. Deux tiers de ces exercices ont été repris dans le produit commercial *lingualevel*; un tiers des exercices a été retenu provisoirement par les cantons pour certains projets d'évaluation.

3.5 Compréhension de l'écrit

La compréhension de l'écrit en langue étrangère se conçoit, dans le cadre d'HarmoS, comme une aptitude langagière de communication, dont l'acquisition peut être considérablement facilitée par l'enseignement disciplinaire des langues étrangères. Les sections suivantes présentent plus en détail les enquêtes qui ont été menées sur la compréhension de l'écrit. Le but poursuivi par ces enquêtes était de constituer une base empirique pour la définition et l'illustration de standards de base en lecture, et de contribuer à concrétiser et à valider, de manière spécifique à cette aptitude, le modèle de compétences «Langues étrangères».

3.5.1 «Construct» et bases conceptuelles

Le *construct* «Compétences en lecture dans une langue étrangère» repose particulièrement sur trois concepts, qui décrivent des aspects essentiels de la lecture, significatifs pour les groupes cibles (concernant le caractère fondamentalement sélectif d'un *construct*, cf. les réflexions présentées en rapport avec le *student model* au point 3.3.1). Ces concepts sont les suivants: a) les processus de la lecture, b) une notion fonctionnelle de la littératie et, partant, des aspects opérationnels de la lecture, ainsi que c) les types de textes. Les éléments a) et b) permettent de tenir compte des spécificités du domaine de compétences «Compréhension de l'écrit»; la catégorie c) constitue une variante, adaptée pour la lecture, des types de discours du modèle de compétences «Langues étrangères», lesquels permettent des différenciations judicieuses dans tous les domaines de compétences. Le *construct* «Compréhension de l'écrit» comporte également, dans une variante spécifiée pour la lecture, les ressources (compétences) retenues dans le modèle de compétences «Langues étrangères» (cf. à ce sujet les développements exposés ci-dessous, au point c).

a) Processus de lecture

Lire est une opération complexe et dynamique, orientée sur la compréhension de textes écrits. La lecture fait intervenir un grand nombre de processus linguistiques et cognitifs, qui interagissent avec différentes sortes de connaissances. On établit souvent une distinction entre les processus qui se produisent au niveau des formes linguistiques et qui se déroulent en grande partie automatiquement chez le lecteur exercé (notamment la reconnaissance lexicale, l'accès au lexique mental, l'analyse de la syntaxe) et les processus permettant la (re)construction de significations ainsi que la compréhension de contenus et de liens logiques, lesquels sont pilotés par des

²⁶ Partie *Ecouter* de la banque de données structurée réunissant des descripteurs, laquelle comprend en tout près de 250 descriptions de compétences spécifiques à des groupes cibles, validées par voie empirique, pour les cinq aptitudes langagières que l'on retrouve aussi dans HarmoS. Les descripteurs sont classés en fonction des niveaux affinés du *Cadre européen de référence*. Au sujet de l'élaboration des descripteurs, se référer à Lenz / Studer (2004).

schémas cognitifs et nécessitent, en règle générale, davantage d'attention (Alderson 2000, Wolff 2002; cf. aussi Nold & Rossa 2006). Les stratégies qui entrent en jeu dans la lecture ont notamment pour fonction essentielle de réguler ces processus en fonction des exigences qui accompagnent une situation de lecture donnée (Phakiti 2007). Cela signifie également que la quantité de processus divers qui interviennent pour la compréhension d'un texte (qu'on appelle souvent «modèle mental» du texte) peut être plus ou moins importante. Ehlers (2006, 35) appelle «souplesse cognitive» («kognitive Flexibilität») la capacité stratégique qui consiste à pouvoir adapter son comportement de lecture aux exigences d'une situation, d'une tâche et d'un texte, et traite dans cette optique les formes de lecture bien connues dans la pratique comme la lecture «sélective» et la lecture «détaillée».

En principe, ces processus sont postulés pour la lecture dans la langue première *comme* dans les langues étrangères. Néanmoins, les processus survenant dans la lecture en langue étrangère sont influencés par différents facteurs, notamment par les connaissances de la langue dont dispose le lecteur (surtout le vocabulaire et la grammaire) et par son aptitude générale à lire, qu'il développe principalement par le biais de la langue première. Les traits typiques d'une lecture en langue étrangère sont une vitesse de lecture moins élevée, mais aussi souvent une capacité réduite à faire des inférences, ce qui, en règle générale, s'explique par des processus encore peu automatisés aux niveaux «inférieurs» de traitement de l'information (par ex. la reconnaissance lexicale) et des connaissances lexicales limitées, ainsi que par des états de savoir qui ne sont que partiellement mobilisables (cf. par ex. Ehlers 2006).

Dans la lignée de la définition générale de la compétence que reprend HarmoS (cf. 3.1), il est particulièrement intéressant pour le *construct* «Compétences en lecture dans une langue étrangère» de voir si les élèves parviennent à adapter leur lecture aux exigences des exercices et des textes qui leur sont proposés: sont-ils par exemple capables de lire un texte attentivement, notamment pour comprendre précisément des explications ou des instructions, et d'en «survoler» un autre pour identifier le thème traité dans le texte? Par rapport à cette question, on établit une distinction générale, en s'inspirant d'Urquhart & Weir 1998, Weir 2005 ainsi que de Sigott 2007, entre la *lecture attentive et précise* («careful reading») d'une part et, de l'autre, la *lecture cursive et sélective* («expeditious reading») ²⁷. La lecture attentive et précise vise une compréhension exacte, dans le détail, et n'opère pas de sélection. Quant à la lecture cursive et sélective, elle comprend plusieurs sortes de lecture dite «rapide», surtout ce qu'on appelle le «skimming» (survoler un texte ou certains passages, par ex. pour identifier le thème d'un texte) et le «scanning» (parcourir un texte pour trouver certaines informations et repérer certains détails).

b) Littérature fonctionnelle et aspects opérationnels de la lecture

Le groupe de travail «Lecture» du consortium Langues étrangères d'HarmoS prend ces «processus» de lecture – avec un accent sur la lecture attentive et précise ainsi que sur la lecture cursive et sélective – comme point de départ, mais les inscrit dans le cadre plus large de ce qu'on peut appeler une «reading literacy», des compétences en lecture. Dans ce cadre, les facteurs qui jouent un rôle majeur sont le contexte social de la lecture, mais aussi les lecteurs eux-mêmes, en tant qu'acteurs sociaux (cf. notamment Urquhart & Weir 1998). Comme la «literacy» implique en règle générale plus que la seule lecture (le fait d'écrire, par exemple, cf. chap. 3.6, mais parfois aussi la «numeracy» [numératie]), il est nécessaire de préciser le terme en spécifiant «*reading literacy*». Dans la perspective des standards de base visés, le groupe de travail comprend par «literacy» une forme de littérature fonctionnelle; c'est donc la question du quotidien écrit des élèves

²⁷ La littérature spécialisée admet un nombre variable de «processus» de ce type, et la terminologie est très peu uniforme dans ce domaine (parmi les termes les plus courants, on trouve, en anglais et en allemand: «operations», «skills», «strategies», «comprehending strategies», «subskills», «Leseformen», «Lesearten» [pratiques de lecture]). Selon Phakiti (2007, 6), il s'agirait d'une stratégie cognitive. Etant donné que le terme «stratégie» évoque parfois de manière assez prononcée une action consciente et du fait que le statut de conscience des formes de lecture n'est pas étudié chez les groupes cibles, nous ne parlons pas ici de stratégies, mais plutôt provisoirement de «processus» (et parfois aussi, en référence à des traditions didactiques, de «Lesearten» [pratiques de lecture]). En psycholinguistique, les processus sont en général considérés comme étant en principe accessibles à la conscience, mais ils peuvent aussi se produire en deçà du seuil d'attention consciente. – N'étudier que deux (groupes de) processus ou pratiques de lecture permet de faire un compromis entre une différenciation motivée au niveau théorique et des nécessités d'opérationnalisation pratiques du *construct*. Le projet spécifique à la lecture des standards de formation autrichiens opère également une distinction entre deux processus de lecture différents (cf. Sigott 2007).

qui se trouve au cœur de la réflexion, de même que celle de savoir comment les élèves se comportent (ou devraient se comporter) dans les différents domaines opérationnels («domains») de leur quotidien écrit. Une littératie comprise dans ce sens comporte deux grands avantages: d'une part, la lecture peut se concevoir, en accord avec le modèle de compétences du consortium Langues étrangères, comme *l'utilisation* de la langue à des fins de communication (et non pas «uniquement» comme un processus psycholinguistique). D'autre part, on se réfère ainsi au concept de «reading literacy» de PISA, un concept qui a certes été développé pour la langue première, mais qui peut aussi s'appliquer à l'apprentissage des langues de manière générale, et peut de ce fait aussi être utilisé, par certains aspects, pour introduire une différenciation dans le *construct* «Compétences en lecture dans une langue étrangère». La définition centrale que donne PISA pour la lecture est la suivante: «Comprendre l'écrit, c'est non seulement comprendre et utiliser des textes écrits, mais aussi réfléchir à leur propos. Cette capacité devrait permettre à chacun de réaliser ses objectifs, de développer ses connaissances et son potentiel et de prendre une part active dans la société» (OCDE, 2003, 120).

Pour affiner la différenciation du *construct*, une autre distinction est établie entre trois aspects opérationnels de la lecture, en référence directe à PISA²⁸: *trouver de l'information, développer une interprétation, réfléchir sur le contenu d'un texte* ou encore *l'évaluer* (OCDE 2003, 2002). Quelques explications à ce sujet: l'aspect *trouver de l'information* consiste à localiser et à comprendre, dans un texte, certaines informations. Pour ce faire, le texte est en général survolé et parcouru par le lecteur, qui cherche à repérer et à identifier les informations qu'il recherche. L'aspect de *l'interprétation* comporte d'abord une compréhension «large», assez générale, de l'ensemble du texte («form a broad general understanding»; par ex. reconnaître les liens présents dans un texte, dans le but d'en comprendre la pensée d'ensemble). Ensuite, dans un second temps, cet aspect engage une compréhension plus exacte ou plus englobante de ce qui a été lu, pour laquelle il est nécessaire de trouver et de comprendre certains liens de corrélation entre des parties du texte ou différents textes (développer une interprétation – «develop an interpretation»). Alors que l'interprétation porte sur des informations internes au texte, l'aspect *réfléchir sur le contenu d'un texte* ou encore *l'évaluer* fait intervenir des connaissances externes au texte qui permettent de faciliter sa compréhension; il s'agit donc de comparer et de relier des informations contenues dans un texte de lecture et des connaissances provenant d'autres sources (par ex. d'autres textes ou ses propres connaissances du monde)²⁹.

c) Types de textes

Nous avons retenu un vaste éventail d'écrits dans le but de couvrir de manière satisfaisante le *construct* «Compétences en lecture dans une langue étrangère», dans la perspective d'une littératie fonctionnelle (cf. chap. 3.2). Le corpus comprend notamment des textes tirés des médias, tout comme des textes relevant du domaine scolaire et privé, des textes narratifs comme des textes informatifs continus et des textes argumentatifs, sans compter des textes non continus. Les «sortes de textes», généralement désignées ainsi («Textsorten») dans la littérature linguistique (surtout Stede 2007), ont été définies de manière plus spécifique pour former des *types de textes*³⁰, sur la base des types de discours du modèle de compétences «Langues étrangères», ainsi que de projets traitant de thématiques comparables (dont le «Dutch CEFR Grid Reading / Listening», cf. www.lancls.ac.uk/fss/projects/grid/).

On distingue cinq types de textes. Les critères appliqués pour opérer cette distinction ont été des caractéristiques linguistiques des textes et des passages de texte (notamment des types de phrases et des signes morphosyntaxiques caractéristiques). Les cinq types retenus sont les

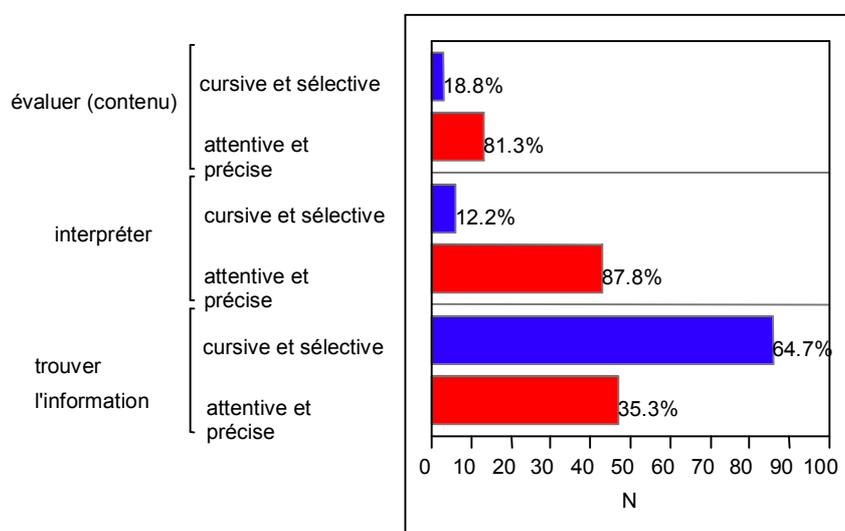
²⁸ Dans PISA, on parle de *processus* ou d'*aspects* de la lecture.

²⁹ Parallèlement à cette forme de réflexion sur les contenus textuels, PISA travaille encore avec une autre catégorie, à savoir le fait de réfléchir sur la forme d'un texte. Cet aspect n'est toutefois pas repris dans le cadre d'HarmoS, puisqu'il ne revêt pas (encore) d'importance particulière pour les groupes cibles dans les disciplines des langues étrangères.

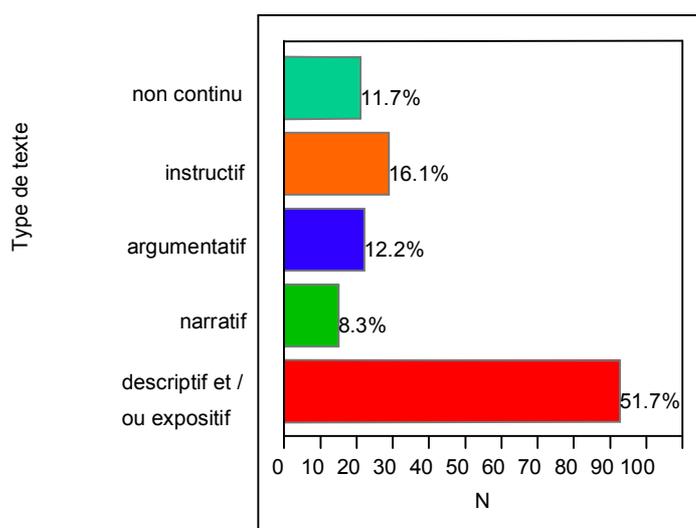
³⁰ Selon Stede (ibid., 47), les types de textes – compris comme des suites de signes linguistiques interprétables du point de vue fonctionnel – peuvent se rapporter à des passages de textes plus ou moins longs, alors que les fonctions textuelles se réfèrent au but poursuivi dans l'ensemble du texte (et accessible au lecteur). Cette terminologie tient compte du fait que de nombreux textes sont composés de passages remplissant des fonctions différentes (narrative, descriptive, etc.). Elle présente l'avantage, pour la codification des textes de lecture, de permettre des *types de textes* différents parmi les passages d'un même texte qui sont nécessaires au lecteur pour traiter les items d'un exercice.

suivants: les textes *descriptifs* et *expositifs* (ces deux types ont été réunis pour des raisons pratiques), *narratifs*, *argumentatifs* et *instructifs*. A cela s'ajoutent les textes *non continus*, c'est-à-dire des textes qui ne sont pas suivis, comme c'est le cas habituellement; il s'agit par ex. de listes d'éléments ou de tableaux. Les textes non continus ont été inclus parce qu'ils revêtent une importance certaine pour les élèves dans la perspective d'une littératie fonctionnelle. Néanmoins, ces textes sont traités comme un groupe à part, du fait qu'ils se caractérisent par une diversité sémiotique bien plus vaste que les types de textes «classiques» et qu'ils doivent de ce fait parfois être décrits autrement. Une question reste également ouverte, celle de savoir dans quelle mesure la compréhension de textes non continus et de textes continus fait intervenir des processus de traitement différents, ou les mêmes, mais avec une importance relative qui peut varier.

Les trois illustrations suivantes représentent la répartition des exercices selon les caractéristiques suivantes: aspects opérationnels et processus de lecture (Ill. 6), de même que type de texte (Ill. 7). Toutes les répartitions se réfèrent au total de 198 items (N=198); ce sont tous des items qui ont pu être gradués (env. quatre items par exercice).



Ill. 6: Répartition des items selon les aspects opérationnels et les processus de lecture



Ill. 7: Répartition des items d'après les types de texte

Ces illustrations montrent que, dans plus de la moitié des items, on a testé une forme de repérage de l'information combinée à une lecture cursive et sélective. Dans près d'un quart des items, l'interprétation ainsi qu'une lecture attentive précise se placent au premier rang. Par comparaison, peu d'items concernent l'aspect «réfléchir sur le texte», dans le sens de l'évaluer. Une grande partie des textes de lecture et des passages revêt un caractère descriptif, mais les textes instructifs et argumentatifs sont aussi bien représentés. Les textes non continus sont représentés par un bon 10%, et les textes narratifs constituent le plus petit groupe.

Dans l'ensemble, le *construct* «Compétences en lecture dans une langue étrangère» comprend donc les composantes suivantes: deux processus ou pratiques de lecture (lecture attentive et précise, ainsi que lecture cursive et sélective), trois aspects opérationnels (trouver de l'information, développer une interprétation et réfléchir sur le contenu d'un texte), de même que cinq types de textes (descriptifs et expositifs, narratifs, argumentatifs et instructifs), avec également le groupe des textes non continus. Ces composantes réunissent les aspects spécifiques de la lecture qui s'avèrent particulièrement importants pour la lecture en tant que compétence d'*utilisation* de la langue, de manière générale, et pour l'acquisition de la lecture des groupes cibles dans les langues étrangères. Dans ces composantes se trouvent aussi déjà intégrées, d'une manière globale, les ressources (compétences) qui sont contenues dans le modèle de compétences «Langues étrangères». Chaque texte de lecture fait appel aux *compétences communicatives langagières* du lecteur (notamment sous la forme de structures simples ou complexes, et de termes fréquents ou moins courants); de plus, faire un exercice de compréhension écrite mobilise aussi toujours des *compétences générales* (par ex. sous la forme de connaissances au sujet du thème abordé dans l'exercice). L'enquête rapportée ici considère de plus près certains aspects de la *compétence linguistique* (en tant que composante significative de la *compétence communicative langagière*), en adoptant la perspective des textes et des exercices. Ces aspects sont présentés ci-dessous, en lien avec des réflexions sur la difficulté des items, parce que l'on part du principe que des caractéristiques linguistiques peuvent influencer la difficulté des items, en conjonction avec d'autres paramètres, comme certaines pratiques de lecture, des aspects opérationnels et certains types de textes (cf. mot clé «item maps» dans la section suivante). Par contre, nous n'avons pas cherché à opérationnaliser les *compétences générales*; nous nous sommes contentés d'effectuer un contrôle, du moins partiel, de ces dernières, par ex. en clarifiant les intérêts thématiques à l'aide des tests préliminaires réalisés, ou en veillant à opter pour différents formats de réponse.

Toutefois, sur la base de la littérature spécialisée en la matière, l'on se montrera prudent, voire sceptique, quant à la question (débatue de manière générale au point 3.3.2) de savoir si les composantes déployées ici et formant le *construct* «Compétences en lecture dans une langue étrangère» peuvent aussi faire l'objet d'une différenciation par voie empirique (cf. Alderson 2000, 95; Weir 2005, 87).

Weir déplace légèrement l'accent dans le débat qui a cours sur la question de la divisibilité ou de la dimensionnalité de la compréhension de l'écrit: selon lui, un maniement minutieux des différents items se révèle plus important que le débat de fond sur la question. Il considère que, quelle que soit la position théorique que l'on adopte, on se met inmanquablement à mesurer des aptitudes / stratégies de lecture (séparément ou par combinaisons) dès que des items distincts sont rédigés sur un passage. Le débat lui semble donc plus porter sur la validité des mises en relation supposées qui cherchent à déterminer quelles aptitudes / stratégies sont testées par quels items, plutôt que d'évaluer la nécessité de traiter la lecture comme une aptitude unifiée³¹.

L'état du débat sur la dimensionnalité de la compréhension de l'écrit a incité le consortium Langues étrangères à prendre deux décisions d'ordre méthodologique:

(1) Puisqu'on ne sait pas précisément si les composantes du *construct* peuvent faire l'objet d'une différenciation empirique, il vaut mieux procéder à des tests qui ne soient pas réducteurs pour pouvoir délimiter une compétence de compréhension écrite assez générale (cf. aussi Alderson 2000, 206 sq.), notamment lorsqu'il s'agit d'une enquête menée à large échelle, comme c'est le

³¹ Citation originale de Weir (2005, 89): *Whatever theoretical position we take, we inevitably measure reading skills / strategies (individually or in combination) as soon as individual items are written on a passage. The debate would seem to be more a question of the strength of any claims concerning which skills / strategies are being tested by which items..., than a strong claim for treating reading as a unified skill.*

cas d'HarmoS. Il peut s'avérer judicieux de réaliser des tests qui visent certains aspects isolés de la compréhension écrite dans le cadre d'études exploratoires de moins grande envergure, par exemple.

(2) Des analyses statistiques portant sur la dimensionnalité du *construct* devraient s'effectuer sur la base d'items décrits en termes précis. Dans le cadre de ce volet de projet, les items sont décrits par rapport aux composantes spécifiques du *construct*, y compris certains aspects participant de la compétence linguistique. Le but premier de ces descriptions est l'interprétation de la difficulté empirique des items.

3.5.2 Instruments

Le *construct* «Compétences en lecture dans une langue étrangère» a été concrétisé, dans un premier temps, par ce qu'on appelle une «**construct map**» (cf. point 3.3.2). Dans le volet de projet portant sur la «Compréhension de l'écrit», cet instrument de travail a d'abord pris la forme de matrices de paramètres échelonnés, puis de descriptions de compétences. Celles-ci se basent d'une part sur les composantes du *construct* qui ont été présentées et, d'autre part, sur des analyses plus globales d'exercices de tests et de descripteurs tirés d'autres projets. Nous avons notamment analysé des exercices relatifs à la compréhension écrite provenant des projets allemands et autrichiens relatifs aux standards de formation³², ainsi que des exercices et des descripteurs du projet alémanique IEF / lingualevel. Nous avons identifié dans ces sources des paramètres qui pouvaient potentiellement déterminer le degré de difficulté (certains significatifs pour le *construct*, d'autres non). Ceux qui sont significatifs pour le *construct* ont été classés en fonction de niveaux provisoires plus larges. Voici par exemple à quoi ressemblait la *construct map* pour la lecture (extrait pour un niveau inférieur «N1») durant une phase initiale du projet, avant l'élaboration des exercices:

N1 (items CÉ IEF)	Caractéristiques des réponses				
	Sorte de texte; paramètres text.	Types de lecture	Egalement significatif pour le <i>construct</i>	Format de la réponse	Pas significatif pour le <i>construct</i> ?
CÉ22, Le17 CÉ 43, Le03 CÉ104, Le01 CÉ23, Le21 CÉ105, Le17	chiffres, suites de mots, phrases simples, textes très courts carton d'invitation; questionnaire simple, panneaux (d'interdiction, d'avertissement); annonces courtes, simples, petites annonces, affiches	Reconnaître, «scanning», attentive-locale: axée sur le «local»	Informations sur la personne; événements et thèmes très connus; mots et tournures très courants, simples	Transfert d'info très simple; correspondance (surtout avec image)	- ressemblance avec l'item en exemple? (CÉ22, item 5) - connaissance du monde? (CÉ104, item 2) - apport de l'image? (CÉ23., item 3: taux de réponse très élevé, item 2: extrêmement difficile) - influence distracteur? (CÉ23, item 5)

III. 8: Extrait de la *construct map* pour la compréhension de l'écrit (phase de travail initiale; les «réponses» sont des reconstructions de paramètres d'items, qui pourraient avoir influencé le comportement des

³² Standards de formation de la Conférence allemande des ministres de l'éducation et des affaires culturelles des länder (KMK) pour la première langue étrangère (anglais / français), destinés au niveau de formation moyen (9^e et 10^e années); décision du 4.12.2003. Standards de formation du Ministère fédéral autrichien de l'éducation, des arts et de la culture (BMUKK) pour l'anglais, destinés aux classes de 8^e année; phase de test jusqu'en 2007. Epreuves communes qui avaient pour but de tester le niveau d'anglais en 9^e année, Rhénanie du Nord - Westphalie, *Ministerium für Schule, Jugend und Kinder*, 2004.

participants au test; CÉ22, etc. = items de compréhension écrite validés, tirés du projet IEF / lingualevel, Le17, etc. = descripteurs validés du même projet)

La *construct map* a été utilisée pour concevoir les exercices (dans le cadre d'un processus itératif, puisque les premières propositions d'exercices ont aussi été mises à profit pour procéder à une différenciation de la *construct map*) et comme ressource pour les descriptions de niveaux définitives.

L'élaboration des **exercices de tests** a été assumée par le groupe de travail «Compréhension de l'écrit» (Verena Tunger, Daniel Elmiger, Thomas Studer). Ce travail peut être résumé en trois étapes. Etape 1: création des exercices; étape 2: traductions; étape 3: phase de tests préliminaires et révision. Le groupe de travail disposait d'environ quatre mois seulement pour réaliser ces trois étapes.

Etape 1: création des exercices

La première étape a consisté à concevoir 21 exercices en tout, sept pour l'allemand, sept pour l'anglais et sept pour le français en langues étrangères, sur la base de la *construct map*, de la littérature spécialisée dans le domaine (par ex. McKay 2006) et des expériences déjà engrangées dans l'élaboration de tests dans le cadre d'autres projets.

Les auteurs ont notamment veillé, dans l'élaboration des propositions, à ce qu'il y ait une répartition relativement équilibrée entre les deux processus de lecture et les trois aspects opérationnels, ainsi que les différents types de textes du *construct*. Ils n'ont toutefois pas cherché à obtenir une transposition totalement égale, dans les exercices, de la structure des composantes du *construct*; en effet, certaines combinaisons de composantes du *construct* s'avèrent plus judicieuses que d'autres, par exemple entre les processus de lecture et les types de textes (Weir 2005, 69): la lecture «scanning» pourra aisément être testée avec des textes descriptifs, parce que ces derniers offrent en règle générale plus de détails factuels, sur lesquels la lecture peut être orientée, que ce n'est le cas habituellement dans des textes argumentatifs. A l'inverse, un tissu de liens qui s'établissent entre les phrases («macropropositions») est plutôt une caractéristique des textes argumentatifs que des textes descriptifs. C'est la raison pour laquelle une lecture cursive («expeditious reading») peut souvent plus facilement être testée avec des textes présentant des opinions que des textes descriptifs. On constate aussi des rapports spécifiques entre des processus et des aspects opérationnels, certains très étroits (par ex. lien évident entre la lecture sélective et la recherche d'informations), d'autres plus vagues (développer une interprétation peut demander une lecture cursive et / ou sélective).

Tous les textes de lectures sont authentiques ou proches de la réalité (par ex. légèrement abrégés). Les principales sources utilisées pour ces textes ont été des revues et des magazines pour les jeunes lecteurs, ainsi qu'Internet (forums, chat, notamment), mais aussi des documents et des prospectus, tirés du domaine scolaire et public, qui sont ou pourraient être intéressants pour les groupes cibles. En prenant en compte différentes sources, il a aussi été possible d'obtenir une certaine diversité de thèmes. Toutefois, les aspects thématiques n'ont pas été systématiquement contrôlés.

Les exercices de tests s'accompagnent en règle générale d'un cadre d'action, c'est-à-dire d'une mise en situation spécifique au groupe cible visé, par ex.: «Tu veux réaménager ta chambre. Sur Internet, tu es tombé-e sur un forum de discussion...», «vous préparez une excursion avec la prof d'anglais...», «tu dois faire un exposé... et tu as lu dans une revue pour les jeunes...». Ces mises en situation dans les exercices de tests ont notamment pour but de rendre plausible le but de la lecture, qui est souvent évident dans des situations hors test. Il s'agit aussi de susciter une certaine motivation à lire, qui est naturelle dans des conditions normales, en dehors de toute situation de test. Pour Alderson (2000, 83), la proximité avec des intentions de lecture survenant dans des situations de vie réelles constitue un critère important de validité pour un test.

Durant la conception des exercices, un soin particulier a été apporté à la sélection du format de réponse. Des résultats récents de la recherche montrent que le format de la réponse peut influencer considérablement la vérification du *construct*. Est particulièrement mis en cause le format du choix multiple (par ex. Rupp / Ferne / Choi 2006, Kobayahi 2002). Ce format pose

problème non seulement parce qu'il s'accompagne de probabilités de deviner la réponse, mais aussi et surtout en raison de l'influence des propositions de réponse qui servent de «distracteurs». En effet, dans la vie réelle, donc en dehors de toute situation de test, la lecture n'est guère associée à des distracteurs. Leur influence peut être plus ou moins grande pour différents participants; elle est de toute manière fondamentalement difficile à contrôler. De l'autre côté du spectre, les formats ouverts soulèvent un autre problème, à savoir que les réponses doivent être interprétées et catégorisées, ce qui peut entraîner des différences considérables dans les évaluations, malgré des clés de résolution et l'entraînement à l'évaluation («rater training»). Trois formats différents ont donc été retenus, de façon à pouvoir au mieux contrôler, tout au moins grossièrement, les effets de distorsion: outre les formats de choix multiples (souvent trois possibilités de réponse), on a eu recours aux formats de comparaison et de classement («multiple matching»), aux questions ouvertes, donc à des réponses courtes, de même qu'à des combinaisons de ces trois formats (par ex. «multiple matching» suivi d'une réponse courte). Alderson (2000, 206) recommande également de combiner les formats; il mentionne aussi les combinaisons de formats dans un seul et même exercice.

C'est dans la langue de scolarisation respective qu'ont été rédigées les mises en situation et les instructions précises permettant de faire les exercices, ainsi que les questions des tests (items). Le choix de ne pas utiliser la langue étrangère pour cela s'explique premièrement par l'intention de s'assurer que les élèves comprennent ce qu'ils doivent faire et, deuxièmement, par le but de tester la compétence en lecture par rapport aux textes à lire, et non pas, par exemple, en lien avec les suggestions d'un exercice qui propose des réponses à choix multiple, réponses qui ont été formulées par les auteurs des tests. Troisièmement, il s'agissait d'éviter que des réponses ne soient données que parce que les mêmes mots apparaissent dans le texte et dans la formulation de l'exercice. Ce ne serait alors qu'un indice de reconnaissance lexicale, et non une preuve fiable de compréhension de la part du lecteur.

Etape 2: traductions

Une sélection de 21 exercices (les originaux) a été traduite dans les autres langues. Les traducteurs devaient rendre le sens du texte (pas de traduction littérale), conserver le caractère du texte et si possible n'apporter aucune modification majeure au texte quant aux difficultés linguistiques qu'il comportait. Les textes traduits ont été contrôlés par des membres du consortium, surtout pour garantir que les items pouvaient aussi clairement être réalisés avec les textes traduits et que leur degré de difficulté ne différait pas trop de celui des textes originaux (des variations qui pourraient être causées par un emploi idiomatique plus marqué de la langue de la part de l'un des traducteurs). Avec cette façon de procéder, il a été possible de créer des exercices «semblables dans leur structure», qui ont pu être utilisés dans toutes les langues cibles. Au final, le corpus de textes, avec les traductions, comprenait 46 exercices de communication. Ensemble, ils couvrent le spectre de niveaux qui comprend tous les groupes cibles testés.

Etape 3: phase de tests préliminaires et révision

Les exercices ont été testés dans 17 classes en tout. Ont pris part à cette phase de test neuf classes de Suisse alémanique et huit classes de Suisse romande (deux classes de 6^e et quatre de 9^e année pour le français, chaque fois trois classes de 6^e et de 9^e pour l'allemand, ainsi que cinq classes de 9^e pour l'anglais).

La phase de test a comporté une vérification quantitative et qualitative des exercices. D'une part, chaque exercice a été présenté à au moins deux, voire trois classes en moyenne. Il a ainsi été possible d'identifier les items qui ne fonctionnaient pas bien (des items caractérisés par un mauvais ajustement au modèle, resp. des valeurs *fit* et discriminatoires insatisfaisantes). D'autre part, les enseignantes et les enseignants ainsi que les élèves ont rempli de courts questionnaires, qui nous ont permis de connaître leur avis sur les exercices et les items. Les enseignantes et les enseignants ont évalué les exercices selon des critères d'intelligibilité, d'attrait, de complexité et en fonction de la durée d'exécution prévue pour chaque exercice. De plus, ils ont consigné les éventuels problèmes rencontrés par les élèves dans leur travail sur les exercices proposés. Quant aux élèves, ils ont été priés de commenter les exercices par rapport à leur intelligibilité et à leur degré de difficulté.

Sur le plan qualitatif, les exercices ont été bien notés, dans l'ensemble. Par contre, un tiers environ des items a dû être révisé sur la base des constatations quantitatives qui ont été faites. Deux exercices ont même été complètement retirés de la série. Trois exercices ont dû être revus, de façon à mieux couvrir le niveau inférieur. En tout, le corpus d'exercices définitif comprend 47 exercices de tests (dont 22 originaux et 25 traductions).

L'illustration suivante montre la répartition des exercices, après révision, en fonction des classes de 6^e et de 9^e année, ainsi que selon les trois langues testées. Le document mis en consultation relatif aux standards de base contient des exemples d'exercices et des séries d'exercices à des fins d'illustration pour les trois langues représentées.

9ème classe anglais (Suisse al. = Suisse R.)													
Pet Fish 15'	Stoneheart 15'		Friends 10'	Papilio 10'		C-Test	Websites 10'		Argent 10'	Vélo 10'		Bee 15'	Chat 15'

9ème classe français (Suisse al.)						9ème classe allemand (Suisse R.)							
Jaël 15'	Moyen Age 15'	Fou 15'	Chat 15'	Popstar 15'	Stoneheart 15'	Stoneheart 15'	Chat 15'	livre 15'	MoyenAge 15'	Footb. 15'	Popstar 15'		
				Argent 10'	Vélo 10'	Argent 10'	Vélo 10'	Chambre 10'	Centre jeun. 10'				

6ème classe français (Suisse al.)						6ème classe allemand (Suisse R.)									
C-Test	Droits d'enf. 10'	Chambre 10'	Friends 10'	Argent 10'	Vélo 10'	Argent 10'	Vélo 10'	chambre 10'	Friends 10'	Droits d'enf. 10'	C-Test				
Anything 10'	Ballo 10'	Steck-br 10'	Websites 10'	Sports d'hiv10'	Eragon 10'	Fou 15'	Footb 15'	Steck-br 10'	Winter sp 10'	Eragon 10'	Anything 10'	Websites 10'	Ballo 10'	Bee 15'	Footb. 15'

1 cellule = 10 minutes

EFL, seulement en 9ème classe: 2x30' plus 2x20'- plus 1x à 20' avec C-Test.

FLE und **DaF** 9ème et 6ème classes: pour chaque langue et chaque classe 3x30' plus 2x 20'

6ème classe: en plus 1 x à 20' avec C-Test .

III. 9: Répartition des exercices relatifs à la compréhension écrite dans l'enquête de validation nationale – couleurs foncées: textes de lecture en langue originale; couleurs claires: textes traduits. Vert: anglais à l'origine; rouge: français à l'origine; bleu: allemand à l'origine; jaune: deux langues de base (exercices reconstitués). Lorsque les titres des exercices sont les mêmes, cela signifie qu'ils sont semblables dans leur structure dans les langues concernées. Les temps indiqués (10' ou 15') correspondent à la durée d'exécution des exercices en minutes.

Dans le but de pouvoir mieux décrire les items, on a développé ce qu'on appelle des *item maps*. Ces *item maps* se composent de paramètres dont on suppose qu'ils ont une influence sur la difficulté des items. Les orientations spécifiques de ces paramètres peuvent être codifiées et l'on peut établir des liens, à l'aide de procédés statistiques, entre la difficulté des items, telle qu'elle a été saisie par voie empirique, et ces paramètres. Ainsi, il devient possible de mieux interpréter (et, partant, de mieux prévoir) la difficulté empirique des items.

Dans le cadre de l'enquête menée sur la compréhension de l'écrit, on a sélectionné en tout 11 paramètres susceptibles de déterminer le degré de difficulté³³, dont les trois composantes ciblées dans le *construct* «Compétences en lecture dans une langue étrangère», le format de réponse des exercices de tests ainsi que sept autres paramètres. Ces autres paramètres portent plus précisément sur la langue utilisée dans les textes de lecture, et particulièrement sur les passages des textes auxquels se rapportent les items. On a ainsi pu aussi mettre davantage en

³³ La sélection s'est surtout opérée sur la base du projet susmentionné «Dutch CEFR Grid Reading / Listening», des études PISA ainsi que de la littérature spécialisée en la matière, dont notamment l'excellent aperçu fourni par Grotjahn (2000).

évidence certaines compétences linguistiques qui doivent intervenir pour que les données d'exercices puissent être exécutées et les textes compris.

Les 11 paramètres ont été définis de manière opérationnelle. Les définitions ont été consignées dans un guide de codification. Chaque item des 47 exercices de tests a reçu un code basé sur ces 11 paramètres.

L'illustration 10 propose un petit extrait des *item maps* qui ont été réalisées; les illustrations 11 et 12 présentent un extrait du guide de codification (l'ill. 11 est un aperçu synoptique, et l'ill. 12 propose un exemple avec le paramètre «processus de lecture» ou «pratique de lecture»); les explications qui suivent résument de quelle manière les orientations des paramètres ont été déterminées (évaluation des experts d'une part et, de l'autre, saisie quantitative).

Item	Aufgabe	Handlungsaspekt	Antwortformat	Texttyp	Leseprozess	Anzahl Wörter	lexikalische Dichte	Lesbarkeit	Wortfrequenz	Textmenge	Grammatik	Lexik
FL311f	Steckbriefe_dF31-1	1	3	1	2	2	1	1	3	2	1	2
FL273f	Jugendzentrum_dF27-3	1	2	1	2	3	1	1	3	1	1	1
FL244f	Taschengeld_dF24-4	2	1	3	1	1	1	1	4	2	2	1

Ill. 10: Extrait des *item maps*. Sont présentés trois items d'exercices pour le français en langue étrangère («Steckbriefe», «Jugendzentrum» et «Taschengeld»). Ces exercices sont reproduits et esquissés à grands traits dans le document mis en consultation relatif aux standards de base. Les paramètres mis en évidence (cellules surlignées en gris) sont exposés plus en détail dans l'ill. 11.

Merkmal	Umschreibung	Merkmalsausprägungen/ Kodes	Charakter der Kodes und Datenniveau	Besonderes; Interpretation der Kodes
Format	Antwortformat der Aufgabe	1= MCQ (Multiple Choice Question) 2= MM (Multiple Matching) 3= offen/kurz (Kurzantwort) 4=Kombi (Kombination von zwei Formaten, z.B. 2 und 3)	Expertenurteil <i>nominal</i>	Gleiche Kodes für alle Items einer Aufgabe
Texttyp	Itembezogene funktionale Haupt-Ausrichtung des Lesetextes	1= deskriptiv und/oder expositorisch 2= narrativ 3= argumentativ 4= instruktiv 5= diskontinuierlich 7= kein (genau spezifizierbarer) Text zur richtigen Lösung des Items	Expertenurteil <i>nominal</i>	Kodes können für Items innerhalb einer Aufgabe verschieden sein
Prozess	Itembezogener dominanter Leseprozess (bzw. Leseart)	1= sorgfältig-genaueres Lesen (careful reading) 2= erkundendes und selektives Lesen (expeditious reading) 7= kein (genau spezifizierbarer) Text zur richtigen Lösung des Items	Expertenurteil <i>nominal</i>	Kodes können für Items innerhalb einer Aufgabe verschieden sein
Lesbarkeit spezifiziert für DaF, FLE und EFL	Lesbarkeitsformeln: Gunning-Fog, Readability, Coleman-Liau	1= erstes Quartil 2= zweites Quartil 3= drittes Quartil 4= viertes Quartil	Quantitativ; Summe der Durchschnittswerte von drei Lesbarkeitsformeln; Kategorisierung nach Quartilen <i>ordinal</i>	Gleiche Werte für alle Items einer Aufgabe The higher, the harder to read
Lexik spezifiziert für DaF, FLE und EFL	Itemrelevanter Anspruch der Lexik des Textes	1= einfach(er)e Lexik 2= komplex(er)e Lexik 7= kein (genau spezifizierbarer) Text zur richtigen Lösung des Items	Expertenurteil <i>ordinal</i>	Werte können für Items innerhalb einer Aufgabe verschieden sein Je höher der Kode, desto höher der Anspruch, desto schwieriger

Ill. 11: Présentation synoptique d'un extrait du guide de codification (brève caractérisation des paramètres «format de réponse», «type de texte», «processus de lecture», «lisibilité» et «lexique»; les trois premiers paramètres ont été codés indépendamment des langues; les deux derniers se rapportent spécifiquement à l'une des langues testées); l'ill. 12 montre, à l'exemple du paramètre «processus de lecture», de quelle manière ces paramètres d'items ont été opérationnalisés.

Merkmal Leseprozess (Leseart) Kodes und Operationalisierungen	Identifizierung des Merkmals Leseprozess (Leseart)
<p>1= Sorgfältig-genaues Lesen (careful reading)</p> <ul style="list-style-type: none"> - Ziel ist genaues Verstehen und Verstehen im Detail. - Typische Beispiele sind: das Identifizieren, Verstehen und Vergleichen von Fakten, Definitionen oder Meinungen; das Verstehen von Hauptgedanken (läuft in der Regel über das Erfassen der Textorganisation und die Gewichtung von Informationen) und/oder von wichtigen Details (z. B. stützenden Argumenten); das Erschließen von impliziten Informationen aufgrund von expliziten Informationen (propositionale Inferenzen). - Das Lesen kann auf Textelemente, auf eine Textstelle oder auf eine grössere Textmenge gerichtet sein. <p><i>Achtung: Sorgfältig-genaues Lesen kann selektives Lesen (vgl. Kode 2) voraussetzen (auch innerhalb von 1-Punkte-Items). Wenn das der Fall ist, wird dennoch der Kode 1 vergeben.</i></p> <p>2= erkundendes und selektives Lesen (expeditious reading)</p> <ul style="list-style-type: none"> - Ziel ist eher rasches, z. T. grobes Verstehen. 	<p>Dieses Merkmal erfasst den Leseprozess (die Leseart), den ein kompetenter Leser beim Lesen des Textes / der Texte wahrscheinlich anwenden würde, um ein Item richtig zu lösen.</p> <p>Leitfrage: Was ist der jeweils erfolgversprechendste Weg / die beste Art, den itemrelevanten Text zu finden und zu verstehen?</p> <p>Konkretisierung der Leitfrage für 1 Pkt.: Welcher Lese-Prozess (welche Leseart) ist von der Konstruktion her intendiert und folglich „angemessen“,</p> <ul style="list-style-type: none"> - um 1-Punkte-Items richtig lösen zu können (d. h. den einen Punkt zu bekommen)? - um 2-Punkte-Items teilweise richtig lösen zu können (d. h. einen von zwei möglichen Punkten zu bekommen)? <p>Wo zutreffend:</p> <p>Konkretisierung der Leitfrage für 2. Pkt.: Welcher Lese-Prozess (welche Leseart) ist von der</p>

III. 12: Opérationnalisation du paramètre «processus de lecture» ou «pratique de lecture» dans le guide de codification (extrait)

Les travaux de codification des items ont été menés parallèlement à l'élaboration des exercices, mais se sont étendus sur une période bien plus longue.

Les différents processus de lecture, les aspects opérationnels et les types de textes, ainsi que le format de réponse, ont été soumis à l'évaluation d'experts (décision par consensus au sein du groupe de travail).

Quatre des sept autres paramètres ont été saisis de manière quantitative. Pour chaque texte de lecture on a calculé: le nombre de mots, la densité lexicale («lexical density», une sorte de «*type-token ratio*», un indice de richesse lexicale), la lisibilité (valeurs moyennes de trois indices de lisibilité: *Gunning-Fog*, «*readability*», *Coleman-Liau*) et la fréquence lexicale (proportion de mots dans un texte qui figurent parmi les 1000 mots les plus courants d'après les dictionnaires des fréquences)³⁴. Ces valeurs peuvent varier selon la langue dans laquelle est rédigé un texte de lecture (originaux / traductions). Ainsi, la base de référence pour tous ces calculs n'a pas été le total de tous les textes du corpus d'exercices, mais plutôt le total des textes de lecture utilisés dans une langue donnée, donc par exemple le total des mots de tous les textes proposés pour l'allemand en langue étrangère («Deutsch als Fremdsprache» - DaF). On a ensuite établi des calculs comparatifs pour voir, par exemple, si un certain texte DaF comportait un grand nombre de mots ou non, par rapport au total de tous les textes DaF.

Les trois autres paramètres ont été saisis au moyen d'évaluations d'experts (en général décision par consensus au sein du groupe de travail). Ils ont estimé: la quantité de texte proposée, le degré de difficulté grammaticale ainsi que la difficulté lexicale. Ces estimations ne se rapportent toutefois pas à l'ensemble du texte de lecture (comme pour les paramètres quantitatifs); elles portent plus spécifiquement sur les passages que les experts considèrent comme étant nécessaires pour trouver les réponses aux items.

Ces items codifiés ont créé une base permettant de clarifier le degré de difficulté des items, et donc également de valider (en partie) le *construct*.

³⁴ Voici les sources utilisées pour ce calcul: pour le nombre de mots, la densité lexicale et la lisibilité: une ressource accessible sur Internet du nom de «*topicalizer*» (www.topicalizer.com) ainsi que, pour la phase de contrôle, le logiciel LingoFox (www.lingofox.de); pour la fréquence lexicale: les données résultant du projet réalisé à l'Université de Leipzig sur le vocabulaire (<http://wortschatz.uni-leipzig.de/html/sitemap.html>) de même que, à des fins de comparaison pour l'anglais et le français, une ressource canadienne proposée sur Internet: «*lexutor*» (www.lexutor.ca).

3.5.3 Collecte des données

Dans l'enquête de validation qui a été menée à l'échelle nationale, les 47 exercices relatifs à la compréhension de l'écrit ont été répartis dans 32 cahiers de tests différents. Les cahiers se composaient de deux à trois exercices. Chaque cahier a été fait par 150 à 250 élèves (pour des indications plus précises concernant la collecte des données, se référer aux tableaux du point 3.2). Un exercice comprenait en moyenne cinq items. Les durées d'exécution prévues pour chaque exercice étaient de 10 ou de 15 minutes. Tous les exercices devaient être faits sans aucune aide extérieure.

Dix exercices ont été proposés tant en 6^e année qu'en 9^e année, dont notamment les exercices «Vélo», «Argent» et «Zimmer» (cf. ill. 9). L'ill. 9 montre aussi les recoupements entre les langues; les exercices «Vélo» et «Argent», conçus pour le français en langue étrangère, ont par exemple aussi été utilisés, sous la forme de textes traduits présentant une structure semblable, pour l'allemand et l'anglais en langues étrangères.

3.5.4 Correction

Quatre étudiants ont été sollicités pour la correction et la saisie électronique des cahiers retournés. Chaque personne a assumé la responsabilité de huit cahiers de tests, ce qui correspond à un volume de plus de 1000 cahiers d'élèves. Les étudiants ont passé entre 35 et 50 heures à effectuer ce travail.

Les cahiers de tests ont été corrigés à l'aide d'une clé de résolution (des exemples de ces clés de résolution ont été inclus, avec les exercices de tests qui servent à illustrer les standards, dans le document mis en consultation). La correction s'est effectuée différemment selon le format de réponse de chaque exercice. Avec les formats fermés, les réponses des élèves n'ont pas été évaluées, mais «enregistrées», c'est-à-dire qu'on a saisi la position d'une réponse (on a par ex. noté la série «1-0-0-0-0» si seul le premier champ d'un tableau contenant cinq colonnes était coché), que la réponse soit juste ou fausse. Avec un tel procédé, la correction est plus facile à saisir et, de plus, le comportement de réponse des élèves peut être analysé de plus près ultérieurement (possibilité d'analyser les éventuels distracteurs). Dans les questions ouvertes, les réponses des élèves ont été évaluées en fonction d'un schéma juste-faux.

Les réponses enregistrées et évaluées ont été saisies par la même personne, à l'aide de masques excel. On a renoncé, notamment pour des raisons financières, à procéder à une codification systématique double. Une mesure d'assurance qualité a consisté à limiter les possibilités de saisie dans ces masques en fonction des exercices et des items. Il a ainsi été possible de limiter considérablement le risque de faire des erreurs de saisie.

Différentes mesures ont été prises afin de parvenir, dans l'ensemble, à une correction aussi fiable que possible des données collectées: formation des correctrices lors d'une journée d'introduction, mise en place d'une aide en ligne («hotline») pour assurer l'accompagnement du travail individuel de correction, contrôles aléatoires par un deuxième correcteur, particulièrement lors de la remise des données. De plus, une expérience de petite envergure a été menée, à des fins de contrôle supplémentaire, avec des cahiers de tests pour l'anglais. Cette expérience a permis de voir si les correctrices appliquaient uniformément les critères et les codes de la clé de résolution (vérification de la fiabilité inter-codeur). Les résultats de cette expérience ont été très positifs³⁵.

De manière générale, on peut partir du principe que les données relatives à la compréhension de l'écrit ont été corrigées et saisies de manière fiable.

3.5.5 Résultats et analyse des résultats

Les sections suivantes traiteront de certains résultats en particulier, qui se rapportent a) aux exercices et aux items des tests, ainsi qu'à l'élaboration des échelles, b) à la question de la dimensionnalité du *construct* «Compétences en lecture dans une langue étrangère» et c) à la

³⁵ Deux correctrices ont corrigé, indépendamment l'une de l'autre, 100 items (dix cahiers d'élèves avec deux exercices comportant chacun cinq items). Il s'agissait d'exercices avec des formats ouverts. Il en a résulté un fort recoupement des évaluations: kappa 0.9369, erreur-type 0.0206.

distribution, au sein de la population, des compétences en lecture dans une langue étrangère. Le point b) abordera également la question des conséquences qu'entraînent certains résultats sur des aspects de spécifications de tests.

a) En tout, 198 items relatifs à la compréhension de l'écrit ont pu être mis en échelle, 86 pour l'allemand, 82 pour le français et 30 pour l'anglais. Les 198 items se répartissent en 47 exercices, ce qui fait environ quatre items par exercice. Initialement, un exercice comportait en moyenne cinq items. La perte d'items lors de la mise en échelle s'élève donc dans l'ensemble à environ 20%. Les critères de sélection statistiques lors de l'élaboration des échelles ont été les suivants: conformité ou «ajustement» des items au modèle (modèle de *Rasch*; *infit* < 1.2) et sélectivité des items suffisante (discrimination > 0.3), de même que, pour une partie des exercices, fonctionnement identique des items en Suisse alémanique et en Suisse romande d'une part et, de l'autre, en 6^e et en 9^e année. Les échelles suivantes ont été constituées sur la base de ces critères: d'abord une échelle pour l'allemand, une pour le français et une pour l'anglais, puis une échelle commune pour l'allemand et le français (sans l'anglais), et enfin l'échelle définitive, qui comprend les trois langues. Au fil de l'élaboration de cette échelle, il est apparu que de nombreux items d'anglais avaient fonctionné autrement en Suisse alémanique qu'en Suisse romande, et qu'il fallait donc les écarter. En revanche, de nombreux items dans des exercices traduits ont fonctionné de la même manière pour l'allemand (en Suisse romande) et pour le français (en Suisse alémanique), ce qui a permis une perte d'items plus faible dans ces deux langues. Ce qui est surprenant, ce n'est pas uniquement la perte considérable d'items en anglais (mêmes exercices dans les deux parties du pays!), mais aussi le parallélisme important d'items pour le français et l'allemand (traductions d'exercices semblables «seulement» dans leur structure). Il n'est pas aisé d'expliquer ces résultats; dans le cas de l'anglais, la réponse pourrait être liée à la différence de langue première des élèves de Suisse alémanique et de Suisse romande, mais aussi à la place différente qui est accordée à l'anglais dans ces deux parties du pays; cependant, ce ne sont là que des hypothèses provisoires, qu'il s'agirait d'étudier de plus près.

Après évaluation par des experts (consensus de trois personnes qui connaissent très bien le *Cadre de référence*, et qui ont aussi pris part à différentes conférences d'évaluation comparative [«benchmarking»]), des liens de correspondance ont été établis entre les échelles d'items et les niveaux du *Cadre européen de référence*, ainsi que les niveaux affinés différenciés du projet *IEF / lingualevel*. Pour ce faire, on est parti d'exercices et d'items dont la difficulté empirique rejoint (largement) le degré de difficulté estimé. Après avoir été classés en fonction de l'échelle de référence, ils ont joué la fonction d'items d'ancrage. Le classement des autres items s'est effectué en grande partie par procédé mathématique, les items «limite» (*cut-offs*) entre les niveaux ayant aussi parfois été fixés à vue d'œil, à des distances à peu près égales.

De cette manière, il a finalement été possible de constituer une échelle commune uniforme, c'est-à-dire valable pour les trois langues et pour les deux niveaux scolaires testés, sur laquelle chaque item correspond à un niveau affiné de l'échelle de référence européenne. De plus, les niveaux HarmoS pour les classes de 8^e et de 11^e années ont également été représentés sur cette échelle. L'échelle comporte des items pour les niveaux européens entre A1.1 et B2, et pour les niveaux HarmoS I₈ à V₁₁, mais le niveau le plus bas ne compte que quelques items seulement.

Pour mieux comprendre la difficulté empirique des items, qui est en partie plausible, en partie difficile à saisir, des thèses ont été débattues au sein du groupe de travail sur la divergence constatée entre la difficulté empirique et le degré de difficulté estimé pour certains items³⁶. De plus, les 11 paramètres susceptibles de déterminer le degré de difficulté, d'après lesquels chaque item a été codifié (cf. 3.5.2, «item maps»), ont été soumis, avec les échelles empiriques, à différentes analyses statistiques. Des procédés de *clustering* ont été employés, ainsi que de partitionnement et d'analyse des correspondances multiples. La question principale était de savoir quelle part jouent séparément et conjointement les différents paramètres (par ex. la quantité de texte, la

³⁶ Les divergences vont dans les deux sens: il y a des items qui sont plus difficiles empiriquement que ce n'était prévu, et d'autres pour lesquels on constate le contraire. Pour le premier cas, certaines thèses portent sur les formats de réponse au sens large (par ex. difficulté accrue par le nombre de possibilités de réponse dans les exercices «multiple matching» ou genre de distracteurs dans les exercices à choix multiples). Parmi les facteurs qui interviennent dans les items qui se sont avérés plus faciles, empiriquement, que prévu, on pourrait notamment trouver des aspects relevant des compétences générales, comme des propriétés linguistiques des textes et des items eux-mêmes.

fréquence lexicale, la difficulté grammaticale) dans les degrés de difficulté constatés de manière empirique. Les résultats de ces analyses révèlent une certaine prédominance des paramètres lexicaux. Néanmoins, ils montrent aussi que tous les paramètres inclus ne peuvent, ensemble, «expliquer» qu'un peu moins de la moitié de la difficulté des items ($R^2 = 0.46^{37}$). Notons toutefois que ce résultat se situe tout à fait dans l'ordre de grandeur des résultats obtenus dans des projets comparables³⁸.

b) Dans les analyses sur la dimensionnalité du *construct* «Compétences en lecture dans une langue étrangère», il s'agit de voir si les différents éléments de compétences (processus de lecture, aspects opérationnels et types de textes) peuvent aussi être différenciés par voie empirique. Des possibilités de différenciation empirique laisseraient à penser que ces éléments agissent différemment selon les élèves. On ne peut apporter qu'une réponse provisoire à cette question sur la base des analyses effectuées pour le français et l'allemand (calcul de corrélations latentes directement lors de la mise en échelle des items avec ConQuest). Les résultats les plus clairs sont ceux qui concernent les processus de lecture: il se pourrait véritablement que la *lecture attentive et précise* et la *lecture cursive et sélective* forment deux dimensions différentes (corrélation latente de 0.72³⁹). Pour les aspects opérationnels, on dispose certes d'indices qui laissent entrevoir une certaine autonomie de l'aspect *réfléchir sur le contenu d'un texte / l'évaluer*, mais cet aspect n'est représenté que par quelques items uniquement. Par contre, il semble que les aspects *trouver de l'information* et *développer une interprétation* se distinguent à peine comme des dimensions distinctes; à noter toutefois que ce constat pourrait aussi être dû à des problèmes d'opérationnalisation (en partie des problèmes de délimitation entre les aspects *trouver de l'information* et *développer une interprétation*)⁴⁰. Enfin, pour les types de textes, l'image que l'on obtient n'est pas uniforme⁴¹. – Pour conclure, on constate qu'il y a quelques indications empiriques qui laissent à penser que la *lecture attentive et précise*, d'un côté, et, de l'autre, la *lecture cursive et sélective* fonctionnent différemment selon les élèves (cf. sur ce point, se référer aussi à Sigott 2007). Ainsi donc, un bon lecteur, attentif, ne serait pas forcément en mesure de bien répondre à des questions nécessitant une lecture cursive et sélective – et inversement. Par contre, les indices empiriques ne sont pas suffisamment clairs, ou trop hétérogènes, pour établir des assertions correspondantes sur les aspects opérationnels et les types de textes.

Comme on pouvait s'y attendre, le format des réponses semble aussi avoir une influence assez importante sur le *construct*. Les résultats des analyses statistiques laissent entrevoir que les trois formats «choix multiple», «multiple matching» et «réponse courte» pourraient former des dimensions distinctes au niveau empirique⁴². Si cette constatation se trouvait confirmée dans le

³⁷ Cette valeur a été calculée par partitionnement (*partitioning*), au moyen du logiciel de statistique JMP, un procédé récursif, comprenant plusieurs étapes, qui adopte une approche «top-down», contrairement au *clustering* hiérarchique, de type «bottom-up», et qui permet de faire des recherches en fonction des critères qui fournissent un groupement optimal des données. Si ce procédé a souligné l'importance des critères lexicaux, il a également mis en évidence l'influence des formats de réponse.

³⁸ Par exemple, dans le projet DESI spécifique à la compréhension de l'écrit en anglais comme langue étrangère, le modèle de régression explique, avec $R^2 = 0.45$, à peine la moitié des différences observées entre les degrés de difficulté des items (analyse de régression linéaire multiple; Nold / Rossa 2006, 205). Toutefois, les résultats du projet DESI et d'HarmoS ne sont comparables que sur certains points, parce que ces projets se distinguent l'un de l'autre par une série de différences d'ordre conceptuel et méthodologique.

³⁹ Les corrélations latentes s'entendent avec correction des erreurs de mesure et peuvent de ce fait être considérées comme des valeurs maximales.

⁴⁰ Corrélation de 0.78 entre *réfléchir sur le contenu d'un texte / l'évaluer* et *trouver de l'information*; corrélation de 0.77 entre *réfléchir sur le contenu d'un texte / l'évaluer* et *développer une interprétation*. Par contre, la corrélation entre *trouver de l'information* et *développer une interprétation* est plus élevée (0.87), ce qui renforce l'idée d'une indivisibilité de ces deux aspects.

⁴¹ Si certains indices donnent à penser que les types de textes *descriptifs* (y compr. *expositifs*) et *instructifs* d'un côté, et les textes *narratifs* et *argumentatifs* de l'autre influent de la même manière sur la difficulté des items (corrélations de 0.80 et plus au sein de ces deux «groupes» de types de textes, mais corrélations plus faibles entre les «groupes», par ex. 0.66 entre les textes *descriptifs* et *narratifs*). Cependant, on constate aussi par ex. une corrélation relativement élevée, de 0.81, entre les textes non continus et les types de textes *descriptifs*. Dans l'ensemble, on ne peut guère, sur la base de ces analyses, parler des types de textes comme constituant des dimensions de la lecture.

⁴² Les valeurs pour les corrélations latentes sont les suivantes, entre le «choix multiple» (CM) et le «multiple matching» (MM): 0.67, entre le «choix multiple» et la «réponse courte»: 0.62 et entre le «choix multiple» et des «formats combinés»: 0.73. Les corrélations entre les trois formats pris séparément (CM, MM, réponse courte) et les combinaisons

cadre d'autres enquêtes, il faudrait partir du principe qu'une élève sachant par exemple bien gérer les exercices avec réponses à choix multiples ne réalisera pas forcément de bonnes performances dans des exercices de «multiple matching» ou des exercices nécessitant des réponses courtes. Se pose toutefois la question de savoir si les formats des exercices constituent uniquement des facteurs qui perturbent la mesure des compétences en rapport avec le *construct*; en effet, le format «multiple matching» pourrait par exemple être plus susceptible de susciter une lecture attentive et comparative que le format non spécifique du choix multiple⁴³. Quoi qu'il en soit, il pourrait être judicieux – s'il s'avère justement que les formats ont une influence sur les résultats dans le sens d'une différenciation des performances – d'utiliser, en vue de promouvoir l'équité dans les tests, différents formats dans une même épreuve; en effet, avec différents formats, les participants qui rencontrent des difficultés avec un certain format de réponse seront moins fortement désavantagés que si *un seul* format est utilisé. A moins que l'on constate que certains formats sont tout simplement meilleurs que d'autres, parce qu'ils gênent moins l'accès aux compétences ciblées dans le *construct*. – Retenons pour l'instant que le format de réponse peut influencer considérablement sur les performances en lecture, un état de fait qui devrait être encore davantage pris en compte pour les futures conceptions de tests, et notamment dans le cadre du monitoring de l'éducation qui est prévu, mais par ex. aussi pour l'élaboration ou la révision de certificats. Il faudrait également en faire l'objet d'enquêtes préliminaires.

Tous les résultats obtenus sur la dimensionnalité de la compréhension de l'écrit ont été limités par le fait que les analyses à ce sujet n'ont pu être suffisamment poursuivies dans le cadre d'HarmoS (on n'a par ex. pas opéré de différenciation en fonction de l'âge et de la langue des groupes cibles).

c) Cette section présente tout d'abord un aperçu sous forme de tableau contenant les résultats obtenus par les élèves de sixième et de neuvième année dans les tests de lecture. Puis ces résultats sont brièvement explicités. Une échelle unique a été utilisée pour représenter les résultats dans les trois langues étrangères. Les valeurs (points) proviennent de la mise en échelle des données des tests. La moyenne de 500 points sur l'échelle correspond à la performance moyenne de tous les élèves qui ont fait des exercices de lecture en allemand ou en français. L'écart-type de la distribution correspondante équivaut à 100 points. L'échelle empirique pour l'anglais a été adaptée à cette échelle ultérieurement. La valeur indiquée pour chaque standard de base est accompagnée d'une référence à l'échelle européenne.

de formats sont particulièrement difficiles à interpréter, du fait que différentes combinaisons sont réunies dans les formats combinés.

⁴³ Cependant, les enquêtes quantitatives ne permettent en principe pas de savoir si ce qui est prévu, dans l'exercice, comme une lecture cursive, par exemple, est aussi réalisé conformément à l'intention des auteurs par les personnes prenant part au test. Il serait possible d'en apprendre davantage sur de telles questions en optant pour des procédés qualitatifs qui incluraient des comptes rendus de réflexions à haute voix.

Exercices de lecture	Français
Suisse alémanique	
6^e année	
Standard de base – 1^{re} LÉ (A2.1)	465
Standard de base – 2^e LÉ (A1.2)	400
6^e année tous	480
Aujourd'hui, le standard de base pour la 1 ^{re} LÉ serait atteint par 55 % des élèves de sixième année.	
Aujourd'hui, le standard de base pour la 2 ^e LÉ serait atteint par 80 % des élèves de sixième année.	
6 ^e année – filles	492
6 ^e année – garçons	471
9^e année	
Standard de base 1^{re} / 2^e LÉ (A2.2)	530
9^e année tous	523
Aujourd'hui, le standard de base serait atteint par 50 % des élèves de neuvième année.	
9 ^e année – filles	537
9 ^e année – garçons	514
9 ^e année – exigences élémentaires	454
9 ^e année – exigences élevées	537
9 ^e année – exigences supérieures	599

Exercices de lecture	Allemand
Suisse romande	
6^e année	
Standard de base – 1^{re} LÉ (A2.1)	465
6^e année tous	487
Aujourd'hui, le standard de base serait atteint par 60 % des élèves de sixième année.	
6 ^e année – filles	496
6 ^e année – garçons	480
9^e année	
Standard de base 1^{re} / 2^e LÉ (A2.2)	530
9^e année tous	534
Aujourd'hui, le standard de base serait atteint par 50 % des élèves de neuvième année.	
9 ^e année – filles	540
9 ^e année – garçons	525
9 ^e année – exigences élémentaires	470
9 ^e année – sans séparation	512
9 ^e année – exigences élevées	530
9 ^e année – exigences supérieures	575

Exercices de lecture	
Suisse alémanique et Suisse romande	Anglais
9^e année	
Standard de base 1^{re} / 2^e LÉ (A2.2)	530
9^e année tous	512
9 ^e année – filles	527
9 ^e année – garçons	497
CH além.: 9^e année tous	515
Aujourd'hui, le standard de base serait atteint par 45 % des élèves de neuvième année.	
CH além.: 9 ^e année – filles	531
CH além.: 9 ^e année – garçons	498
CH além.: 9 ^e année – exigences élémentaires	451
CH além.: 9 ^e année – exigences élevées	524
CH além.: 9 ^e année – exigences supérieures	593
CH rom.: 9^e année tous	503
Aujourd'hui, le standard de base serait atteint par 40 % des élèves de neuvième année.	
CH rom.: 9 ^e année – filles	513
CH rom.: 9 ^e année – garçons	491
CH rom.: 9 ^e année – exigences élémentaires	414
CH rom.: 9 ^e année – sans séparation	477
CH rom.: 9 ^e année – exigences élevées	497
CH rom.: 9 ^e année – exigences supérieures	558

Certains de ces résultats sont brièvement développés et évalués ci-après⁴⁴. Grâce à l'échelle commune établie, il est possible de comparer entre eux les résultats des tests dans les deux parties du pays et les trois langues étrangères concernées.

Une première comparaison montre que les compétences en lecture des élèves romands en allemand sont légèrement plus prononcées que celles des élèves alémaniques en français, mais la différence n'est pas significative⁴⁵.

Pour le **français en Suisse alémanique**, on obtient l'image suivante: dans les classes de 9^e année, les compétences en lecture sont, comme on pouvait le penser, plus développées que dans les classes de 6^e année⁴⁶. On constate des différences liées au sexe, toujours dans le sens qui

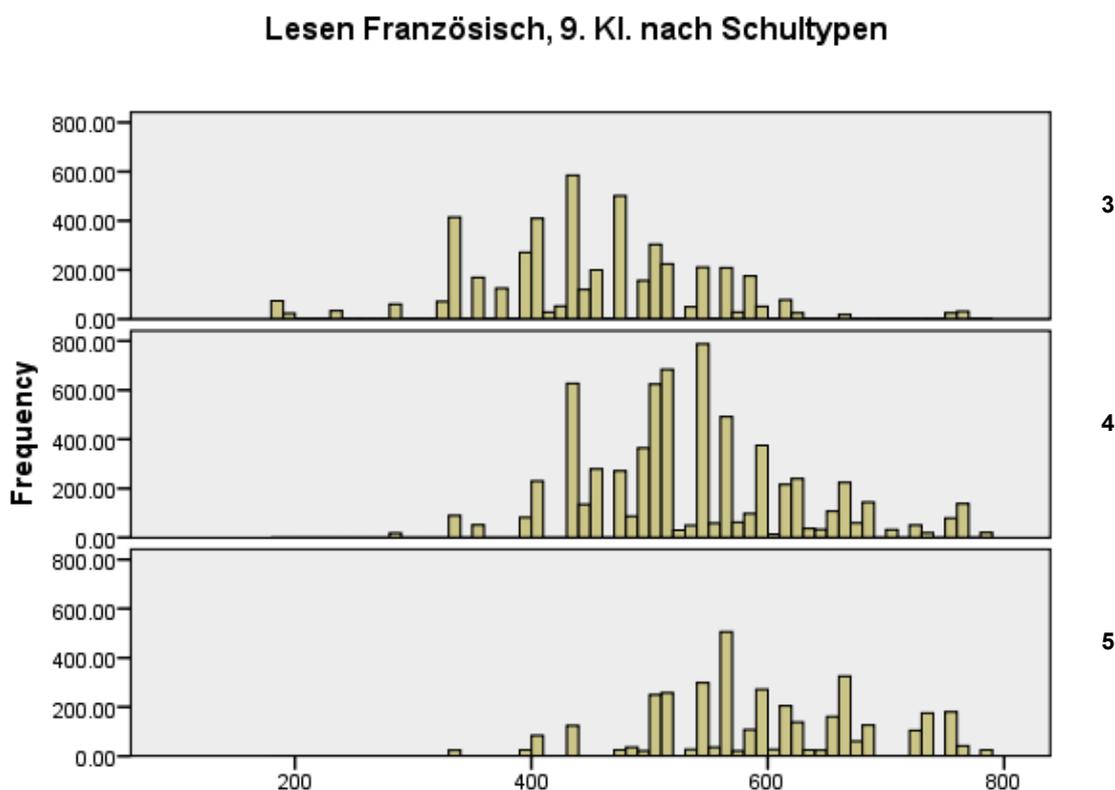
⁴⁴ Des comparaisons entre les groupes de population ont été calculées à l'aide de t-tests et d'analyses de la variance. La signification des résultats est indiquée par le signe *, à savoir que * = significatif (p < 0.05), ** = très significatif (p < 0.01) et *** = extrêmement significatif (p < 0.001). Les données numériques dans les histogrammes ne correspondent pas au nombre d'élèves représentés dans l'échantillon; il s'agit en fait de valeurs extrapolées à la population des élèves. La qualité de l'échantillon a été prise en compte, dans le calcul du niveau de signification, sous la forme d'un effet de plan («design effect») estimé à 2.

⁴⁵ Valeur moyenne pour la Suisse romande: 505, valeur moyenne pour la Suisse alémanique: 499, t=1.016.

⁴⁶ Différence de valeur moyenne entre la 6^e et la 9^e année: 42, t=5.593***

était prévisible, c'est-à-dire avec des avantages pour les filles, mais ces différences ne sont significatives qu'en 6^e année. En 9^e année, les différences ne sont pas (plus) significatives⁴⁷.

L'illustration suivante représente la distribution, en 9^e année, des compétences en lecture selon les types de sections:



Ill. 13: Compétences en lecture pour le français comme langue étrangère en 9^e année (Suisse alémanique) selon les types de sections; 3= exigences élémentaires, 4= exigences élevées, 5= exigences supérieures

L'ill. 13 révèle des différences très importantes entre les classes avec exigences élémentaires d'un côté et, de l'autre, les classes avec exigences élevées et celles avec exigences supérieures⁴⁸. La valeur moyenne dans les classes de 9^e année avec exigences élémentaires est plus faible que la valeur moyenne des classes de 6^e année (454 points contre 480 points).

Les résultats correspondants pour l'**allemand comme langue étrangère en Suisse romande** sont les suivants: on retrouve une différence prévisible entre la 6^e et la 9^e année⁴⁹. Quant aux différences liées au sexe, elles ressemblent également passablement aux résultats obtenus pour le français en Suisse alémanique. On constate là aussi que les compétences en lecture sont plus élevées chez les filles que chez les garçons, mais ces différences sont (à peine) significatives en 6^e année seulement. En 9^e année, les différences ne sont pas (plus) significatives⁵⁰.

L'illustration suivante montre la distribution, en 9^e année, des compétences en lecture selon les types de sections:

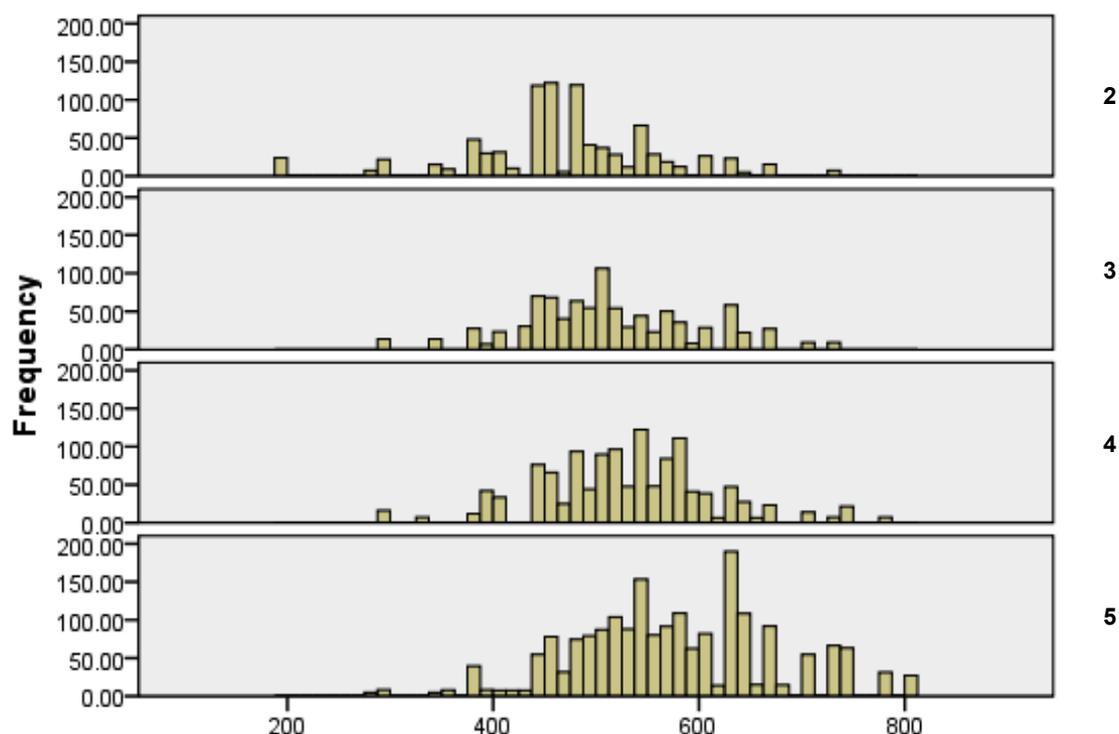
⁴⁷ Différences de valeur moyenne entre les filles et les garçons, 6^e année: 21, $t=2.323^*$; 9^e année: 23, $t=1.675$.

⁴⁸ La différence entre les valeurs moyennes s'élève à 105 points, $t= 7.894^{***}$; une analyse de la variance incluant les trois types de classes confirme cette différence importante: $F= 42.114^{***}$, $df=2$.

⁴⁹ Différence de valeur moyenne entre la 6^e et la 9^e année: 46, $t=6.670^{***}$

⁵⁰ Différences de valeur moyenne entre les filles et les garçons, 6^e année: 16, $t=1.963^*$; 9^e année: 15, $t=1.422$

Lesen Deutsch, 9. Kl. nach Schultypen



Ill. 14: Compétences en lecture pour l'allemand comme langue étrangère en 9^e année (Suisse romande) selon les types de sections; 2= exigences élémentaires, 3= sans séparation, 4= exigences élevées, 5= exigences supérieures

L'ill. 14 montre que les différences entre les divers types de sections en Suisse romande doivent être évaluées de manière différenciée. On observe des différences significatives entre les deux sections inférieures («exigences élémentaires» et «sans séparation») ainsi qu'entre les deux types de sections supérieures («exigences élevées» et «exigences supérieures»). Par contre, les différences entre les deux sections moyennes ne sont pas significatives, donc entre les classes intégrées et les classes avec des exigences accrues («sans séparation» et «exigences élevées») ⁵¹. En Suisse romande aussi, on remarque que les compétences en lecture dans les classes de 6^e année se profilent en moyenne mieux que dans le niveau inférieur de 9^e année (487 points contre 470 points).

Pour l'**anglais comme langue étrangère**, les classes de 9^e année atteignent dans l'ensemble une valeur moyenne de 512 points. En comparaison des valeurs moyennes pour l'allemand (533 points) et le français (523 points) en 9^e année, cette valeur est légèrement inférieure. Cette différence est minime si l'on songe que le temps d'apprentissage moyen pour l'anglais avait été bien moins long, au moment où les données ont été collectées, pour l'anglais que pour les langues nationales ⁵². En Suisse alémanique, la valeur moyenne pour l'anglais est légèrement supérieure à celle en Suisse romande, mais la différence n'est pas significative ⁵³. On constate une différence de

⁵¹ Différence de valeur moyenne entre les «exigences élémentaires» et le type «sans séparation»: 42, $t=2.666^{**}$; différence de valeur moyenne entre les «exigences élevées» et les «exigences supérieures»: 45, $t=3.607^{***}$; différence de valeur moyenne entre le type «sans séparation» et les «exigences élevées»: 18, $t=1.275$. Dans l'ensemble, c'est-à-dire si l'on prend en compte les quatre types de classes, la différence est hautement significative ($F=20.01^{***}$, $df=3$).

⁵² Il s'agit toutefois d'interpréter avec prudence les comparaisons établies entre l'allemand et le français d'un côté et, de l'autre, l'anglais, parce que l'échelle pour l'anglais n'est pas directement rattachée par des moyens statistiques à l'échelle pour l'allemand et le français; la mise en regard a en fait été réalisée ultérieurement, par le biais d'un procédé qualitatif.

⁵³ Différence de valeur moyenne: 13, $t=1.177$

maîtrise de l'anglais entre les filles et les garçons dans les deux parties du pays, mais la différence n'est significative qu'en Suisse alémanique⁵⁴. Par contre, on observe des différences notables en anglais entre les types de sections, tant en Suisse alémanique qu'en Suisse romande. En Suisse alémanique, la différence n'est pas seulement significative entre la section inférieure (exigences élémentaires) et les sections supérieures (exigences élevées et exigences supérieures); elle est aussi marquée entre les exigences élevées et les exigences supérieures⁵⁵. En Suisse romande, la situation pour l'anglais est identique à celle pour l'allemand: les différences sont significatives entre les deux sections inférieures («exigences élémentaires» et «sans séparation») ainsi qu'entre les deux sections supérieures («exigences élevées» et «exigences supérieures»). En revanche, les différences constatées ne sont pas significatives entre les deux types de sections moyennes («sans séparation» et «exigences élevées»)⁵⁶.

On peut dresser le bilan suivant à partir des résultats relatifs à la population: les compétences en lecture dans les langues étrangères suivent une distribution similaire en Suisse alémanique et en Suisse romande. Des différences très marquées se font jour, dans les deux parties du pays, entre les différents types de sections, et ce tant pour les langues nationales que pour l'anglais. Citons encore un élément qui retient l'attention: les compétences en lecture sont, en moyenne, plus prononcées dans les classes de 6^e année que dans le niveau inférieur des classes de 9^e année. Cette constatation est particulièrement significative pour la Suisse alémanique, parce qu'une plus grande proportion d'élèves de 9^e année y suit sa scolarité dans des classes de type «exigences élémentaires». Si les différences liées au sexe sont particulièrement marquées en 6^e année, elles semblent toutefois s'atténuer jusqu'en 9^e année.

3.5.6 Bilan et perspectives

Il a été possible de constituer, pour la compréhension de l'écrit en allemand, en français et en anglais comme langues étrangères, une échelle uniforme sur laquelle on a pu situer les compétences en lecture dont disposaient des élèves de classes de 6^e et de 9^e année au moment de l'enquête. L'échelle comprend près de 200 items (tirés d'environ 50 exercices de tests), lesquels ont tous été attribués à un niveau affiné de l'échelle de référence européenne et sont décrits plus précisément par des caractéristiques qui se rapportent au modèle de compétences «Langues étrangères» et, plus spécifiquement, au *construct* «Compétences en lecture dans une langue étrangère». Ce matériel peut être exploité à diverses fins, notamment pour illustrer, de manière compréhensible, probante et orientée vers le développement, ce qu'on peut attendre, en compétences de lecture, des futurs élèves des classes de 8^e et de 11^e année (cf. illustration du domaine de compétences «Compréhension de l'écrit» et exemples d'exercices de lecture dans le document mis en consultation relatif aux standards de base). Il a ainsi été possible de réaliser l'un des principaux objectifs du travail délimité par le mandat confié.

L'analyse statistique des données a livré quelques résultats importants, tant sur le *construct* «Compétences en lecture dans une langue étrangère» que sur la distribution, au sein de la population, des compétences en lecture.

Les analyses relatives au *construct* peuvent être interprétées ainsi: les deux (groupes de) processus ou pratiques de lecture différents que sont la lecture attentive et précise d'une part et, de l'autre, la lecture cursive et sélective constituent des dimensions autonomes de la lecture. Par contre, les indices empiriques pour les différents aspects opérationnels et les types de textes devraient être plus prononcés et homogènes pour que l'on puisse parler de «dimensions». Les analyses sur la dimensionnalité ont permis de fournir une contribution en vue de la validation empirique du *construct*, même si elles devraient encore faire l'objet d'une différenciation et d'une

⁵⁴ Différence de valeur moyenne pour l'anglais en Suisse alémanique: 33, $t=2.430^*$; différence de valeur moyenne pour l'anglais en Suisse romande: 21, $t=1.743$

⁵⁵ Différence de valeur moyenne entre les exigences élémentaires d'un côté et, de l'autre, les exigences élevées et supérieures: 99, $t=7.502^{***}$; différence de valeur moyenne entre les exigences élevées et les exigences supérieures: 69, $t=4.684^{***}$

⁵⁶ Différence de valeur moyenne entre les «exigences élémentaires» et le type «sans séparation»: 63, $t=3.501^{**}$; différence de valeur moyenne entre les «exigences élevées» et les «exigences supérieures»: 61, $t=4.378^{***}$; différence de valeur moyenne entre le type «sans séparation» et les «exigences élevées»: 19, $t=1.198$.

confirmation supplémentaire, sans quoi il n'est pas possible d'en déduire des conséquences didactiques fiables, par exemple.

Les analyses relatives à la population révèlent notamment des différences notables dans les compétences en lecture selon le type de section. Si l'on veut – comme le prévoit le concept des standards *de base* – qu'à l'avenir les élèves des classes avec exigences élémentaires puissent aussi atteindre les standards de base, il s'avère indispensable que toutes les personnes impliquées dans l'enseignement et l'apprentissage des langues déploient des efforts considérables pour y parvenir.

De plus, les analyses ont permis d'obtenir quelques indications sur le fonctionnement des tests et sur les conditions attachées aux exercices de tests. Parmi les principaux constats obtenus dans ce domaine, on peut citer le fait que les formats de réponse, tels que le choix multiple, le «multiple matching» ou la réponse courte, peuvent exercer une influence considérable sur les résultats, au point d'avoir un effet de différenciation sur les performances réalisées. Cette constatation peut notamment inciter à vérifier les dispositifs de tests existants et à opérer des différenciations dans les futurs tests qui seront élaborés. Etant donné que, pour déterminer les compétences en lecture, on ne peut pas renoncer à des exercices de tests comportant certains formats de réponse, il s'agira de relever le défi, dans le cadre de tests futurs, d'étudier l'influence des formats, de la révéler et de la contrôler aussi bien que possible.

Ajoutons pour terminer que toute une série de travaux complémentaires prometteurs s'est profilée au cours des enquêtes qui ont été réalisées sur la compréhension de l'écrit. Parmi les travaux qui pourraient s'appuyer sur les données collectées, on trouve notamment des analyses comparatives sur

- les bons et les mauvais lecteurs;
- les compétences en lecture des élèves qui, dans le cadre de l'enquête de validation, ont travaillé sur des cahiers de lecture dans deux langues étrangères;
- les compétences en lecture et en écriture dans une langue étrangère, ou les compétences en lecture et les résultats des tests C;
- les compétences en lecture dans la langue étrangère et dans la langue scolaire locale.

Parmi les autres intérêts de recherche, on pourrait mentionner:

- élaboration et développement du catalogue de paramètres qui ont permis de codifier les items des tests: on peut supposer que la «variance expliquée», établie sur la base des différentes caractéristiques, serait plus élevée si l'on intégrait dans le catalogue d'autres paramètres encore, dont le thème des textes ou l'organisation textuelle par exemple, et / ou des prévisions concernant l'intelligibilité des textes de lecture en fonction du contenu traité dans le texte (par ex. sur la base du modèle dit «de Hambourg» («Hamburger Modell»)).
- développement et test d'exercices complémentaires, notamment sur les éléments de compétences qui sont plutôt sous-représentés dans le corpus. Cela implique par ex. des exercices sur l'aspect opérationnel *réfléchir sur le contenu d'un texte* ou des exercices portant sur des textes narratifs. Les résultats qui se dégageraient de telles enquêtes pourraient contribuer à une meilleure évaluation d'indices à disposition sur la «divisibilité» de la compréhension de l'écrit.
- des enquêtes qualitatives portant sur la question du comportement qu'adoptent, très concrètement, les groupes cibles lorsqu'ils sont confrontés aux exercices relatifs à la compréhension de l'écrit. Des études de ce type, qui étaient d'ailleurs prévues dans le cadre d'HarmoS, mais qui n'ont pas pu être menées, pourraient (en complément, mais pas seulement, des analyses quantitatives sur la dimensionnalité de la compréhension de l'écrit) livrer des informations plus précises permettant de savoir si les apprenants adaptent leur comportement de lecture aux exigences de la situation, des textes et des exercices, et s'ils le font comme cela est prévu dans la donnée (par ex. lecture sélective ou lecture précise). Parmi les méthodes qui s'avéreraient appropriées pour de telles études, on trouverait notamment des formes d'observation et de réflexion à haute voix⁵⁷.

⁵⁷ Un premier pas dans cette direction a été entrepris, dans le cadre d'un séminaire au Département des Sciences du plurilinguisme et des langues étrangères de l'Université de Fribourg. Lors de ce séminaire, des élèves ont été interrogés, dans le cadre de projets empiriques de petite envergure, sur les parcours de résolution qu'ils adoptaient pour répondre à des exercices de tests HarMoS relatifs à la compréhension de l'écrit. Les premiers résultats qui se dégagent des

3.5.7 Résumé

La compréhension de l'écrit en langue étrangère a été intégrée, dans le cadre d'HarmoS, comme une aptitude langagière de communication et étudiée par voie empirique lors de l'enquête de validation qui a été menée à l'échelon national. Cette enquête visait à saisir les compétences en lecture (en allemand, en français et en anglais comme langues étrangères) des élèves dans les classes actuelles de 6^e et de 9^e année. Il s'agissait d'une part de créer une base empirique pour déterminer et illustrer des standards de base relatifs à la compréhension de l'écrit et, d'autre part, de contribuer à la concrétisation, par aptitude, et à la validation du modèle de compétences «Langues étrangères».

L'étude a porté en particulier sur certains aspects de la lecture qui s'avèrent significatifs pour l'acquisition de la lecture des groupes cibles dans les langues étrangères et pour l'enseignement disciplinaire des langues étrangères dans la lecture. Les principales thématiques abordées ont été les suivantes: les élèves parviennent-ils à adapter la lecture aux exigences qui se posent dans certaines situations, certains exercices et textes? Autrement dit, sont-ils capables de lire par ex. *un* texte avec attention et précision, et un autre de manière sélective? Avec quelle facilité les élèves réussissent-ils à trouver de l'information dans différents textes? Avec quelle facilité interprètent-ils des textes et réfléchissent-ils sur le contenu d'un texte? Quelle est l'influence de la sorte / fonction d'un texte et de certains aspects qualitatifs des textes de lecture (par ex. difficulté du lexique et de la grammaire, fréquence lexicale) sur le comportement de lecture des élèves?

Ces questions ont été traitées à l'aide d'un corpus de près de 50 exercices de tests axés sur la communication.

Parmi les principaux résultats, on peut nommer les suivants: il a été possible de constituer, pour la compréhension de l'écrit en allemand, en français et en anglais comme langues étrangères, une échelle uniforme, comprenant près de 200 items qui ont tous été classés en fonction des niveaux du *Cadre européen de référence* et des niveaux affinis d'*IEF / lingualevel*. On a pu situer sur cette échelle les compétences en lecture dont disposaient des élèves de 6^e et de 9^e année au moment de l'enquête. Il a ainsi été possible de réaliser l'un des principaux objectifs de l'enquête.

De plus, il a été possible d'apporter une contribution à la validation empirique du *construct* «Compétences en lecture dans une langue étrangère». S'il est clair que certains résultats obtenus dans ce domaine devraient être soumis à des analyses et à des interprétations synoptiques supplémentaires, il est néanmoins apparu que les deux processus ou pratiques de lecture – la lecture attentive et précise d'un côté et, de l'autre, la lecture cursive et sélective – constituent des dimensions que l'on peut distinguer par voie empirique. C'est donc avec raison qu'elles ont été intégrées comme des composantes indépendantes dans le *construct* (selon toute probabilité, tel n'est pas le cas pour les *aspects opérationnels* et les *types de textes*).

En ce qui concerne la distribution des compétences en lecture au sein de la population, on constate qu'elles sont développées à peu près de la même manière chez les élèves de Suisse romande et de Suisse alémanique. De très grandes différences de performances apparaissent toutefois dans les classes de 9^e année en fonction des types de sections, et ce dans les deux parties du pays, tant pour les langues nationales que pour l'anglais.

Enfin, on a pu dégager des informations sur l'effet de différenciation qu'exercent les formats de réponse sur les performances des élèves, des conclusions qui peuvent être exploitées pour de futures spécifications de tests et d'exercices.

réponses obtenues indiquent qu'il y a effectivement des raisons de conclure à un comportement «conforme aux attentes» (par ex. lorsque, dans des exercices sur la lecture sélective, l'élève lit en effet d'abord la question, l'item, et parcourt ensuite le texte pour trouver la réponse). Mais certains signes donnent aussi à penser que ce comportement peut très rapidement être perturbé et qu'il peut être influencé par des facteurs très divers, notamment par des aspects liés au format de la réponse, tout comme des paramètres linguistiques des textes, des items et des distracteurs.

3.6 *Ecrire*

3.6.1 *Situation initiale*

A l'heure actuelle, l'écrit ne revêt pas une grande importance, en tant que domaine de compétences propre, ni dans les manuels scolaires ni dans la pratique de l'enseignement à l'école primaire. On «note» (vocabulaire, exercices), plutôt que de rédiger en vue d'accomplir un acte de communication. Au degré secondaire, cette fonction de soutien de l'écrit demeure, mais les élèves doivent aussi écrire des textes – en général assez courts – dans un but de communication.

L'écrit à des fins de communication trouve un ancrage solide dans les futurs scénarios d'enseignement, dans lesquels des éléments de la pédagogie des échanges ainsi que le recours à des contenus disciplinaires (par ex. dans le cadre du CLIL) jouent un plus grand rôle.

Dans le cadre d'HarmoS, il a été décidé, vu cette évolution, de tester le domaine de compétences de l'écrit, tant en sixième qu'en neuvième année. L'un des principaux enjeux était la perspective de dégager des informations, à l'aide d'exercices d'écriture, sur la capacité des élèves de communiquer par écrit – capacité qui n'est certes guère développée de manière ciblée dans l'enseignement (des langues étrangères), mais qui, selon toute probabilité, est mobilisable, jusqu'à un certain point, chez les élèves. On a considéré que les élèves pourraient transférer, dans le contexte de la langue étrangère, des compétences relevant du domaine de l'écrit axé sur l'action, à partir de l'enseignement des langues reçu à l'école ou encore d'expériences acquises dans leur quotidien. Cependant, il est clair que l'évolution de l'écriture ne peut être considérée comme étant achevée chez la grande majorité des élèves (particulièrement pas chez les élèves de sixième année), même dans la langue L1, que ce soit d'un point de vue cognitif ou sur le plan fonctionnel (compétence relative aux sortes de textes!)⁵⁸.

La base sur laquelle reposent les tests HarmoS, orientés sur l'action, est constituée d'un modèle sociocognitif de l'écrit, qui s'intéresse d'une part à des éléments de compétences linguistico-cognitifs et, de l'autre, à la capacité des apprenants d'intégrer, de manière appropriée, dans leurs textes écrits, des facteurs contextuels, qui résultent de la donnée et du contexte de l'exercice (cf. Shaw / Weir 2007, 62). C'est pourquoi tous les exercices sont assortis d'une mise en situation claire et d'un accent opérationnel bien défini, que les élèves devaient prendre en compte. L'évaluation a porté tant sur l'exécution de l'exercice du point de vue fonctionnel, du contenu, que sur la réalisation linguistique et communicationnelle.

La capacité des élèves de communiquer par écrit a été testée à l'aide de neuf exercices différents. Ces exercices ont été conçus de façon à tenir compte des divers types de discours mentionnés ci-dessus. La distinction qu'opère le *Cadre de référence* entre l'interaction écrite et la production écrite n'a pas, contrairement à l'expression orale (voir ci-dessous), été mise en exergue dans HarmoS; toutefois les tests comportent des exercices sur les deux modes de communication.

Les tests sur l'écrit ont pu être réalisés, au printemps 2007, en complément des tests (de réception) sur la compréhension de l'écrit, dans le cadre de l'enquête représentative, tandis que pour les tests relatifs à l'expression orale, il n'aurait pas été envisageable, pour des raisons pratiques, de faire participer un nombre d'élèves qui soit représentatif.

3.6.2 *Instruments (exercices)*

Les concepteurs des exercices et des instruments d'évaluation – qui se conçoivent comme des moyens complémentaires nécessaires à la mise en œuvre d'un *construct* – se sont notamment inspirés, pour la base conceptuelle du projet, du *Cadre de référence* et des sources déjà mentionnées plus haut, qui ont été réalisées dans cette même lignée, notamment les descriptions de compétences, les critères d'évaluation et les exercices de communication du projet alémanique *IEF / lingualevel*; ils ont également consulté des documents de référence développés par ALTE (2003, a; b) pour l'élaboration et la description d'exercices d'écriture, lesquels ont été inclus dans le projet de manuel du Conseil de l'Europe (2003). Par ailleurs, les auteurs ont eu recours à la

⁵⁸ Cf. notamment pour la compétence en écriture et le *construct* relatif à l'écriture: Baer et al. 1997; Bereiter / Scardamania 1987; Grabe / Kaplan 1996; Harsch 2005; Hayes / Flower 1980 et Shaw / Weir 2007.

littérature spécialisée sur la compétence en écriture, de même qu'à celle relative aux tests et à l'évaluation dans le domaine de l'écriture⁵⁹.

Plusieurs générations de *construct maps* ont été élaborées pour un usage interne, dans le but de servir de base de planification pour les paramètres des exercices et pour les instruments d'évaluation correspondants. Elles comportaient notamment des paramètres issus de descriptions de compétences orientées sur l'action et de grilles d'évaluation spécifiques aux ressources langagières; à cela se sont ajoutés des éléments systématiques, dont le rôle était de garantir que le *construct* soit couvert, en particulier la représentation, dans la série d'exercices, de types de discours jugés essentiels.

Le tableau suivant donne un aperçu des exercices proposés. Il comprend des indications sur les types de discours visés par les (différentes parties d')exercices, ainsi que sur la répartition des exercices en fonction des niveaux des classes.

Type de discours	Exercices d'écriture								
	602 - Envoyer une carte d'anniversaire	801 - Se présenter brièvement	805 - Rendez-vous pour une visite (Tomek)	807 - Un jour d'école ordinaire	808 - Expérience avec un animal domestique	811 - Avis sur les devoirs	906 - Demander des informations sur le chocolat	909 - Un jour inoubliable	912 - Interdiction des téléphones portables
Informier / décrire	X	X		X		X			
Inviter / motiver à faire quelque chose			X				X		
Raconter / rapporter	X				X			X	
Exprimer son opinion / argumenter						X	X	X	X
Maintenir la relation	X		X				X		
Proposés dans les classes suivantes:									
Allemand, 6 ^e année									
Allemand, 9 ^e année									
Français, 6 ^e année									
Français, 9 ^e année									
Anglais, 9 ^e année, CH além. et CH rom.									

Dans la répartition des exercices en fonction des classes, on n'enregistre que quelques différences minimes entre les langues. De nombreux résultats peuvent par conséquent faire l'objet de comparaisons directes dans le cadre d'enquêtes plus poussées.

Selon les exercices, il fallait travailler avec un ou plusieurs type(s) de discours pour répondre à ce qui était demandé. Dans l'exercice 909 («Un jour inoubliable»), par exemple, les lignes directrices demandaient à l'élève de *raconter* un jour inoubliable et d'*expliquer (justifier, argumenter)* pourquoi ce jour avait été particulier.

⁵⁹ A noter en particulier les ouvrages mentionnés dans la note de bas de page précédente; pour des informations spécifiques aux tests et à l'évaluation, cf.: Beck / Klieme (2007); Cushing Weigle 2002; Harsch (2006); McKay (2006); Wolfe-Quintero / Inagaki & Kim (1998).

La présentation des exercices d'écriture dans les cahiers de tests suit, dans les grandes lignes, la structure des exercices que l'on retrouve dans tous les domaines de compétences testés. Concrètement, les exercices d'écriture ressemblaient à ceci (exemple avec l'exercice 906):

1. Instructions générales avec indication du temps à disposition. Les 20 minutes imparties (30 minutes dans le cas de l'exercice 912) étaient calculées assez juste; c'est pourquoi les instructions précisaient que les élèves ne devaient pas avoir trop peur de faire des fautes. Cette précision a été ajoutée après quelque hésitation, parce qu'elle pouvait être mal interprétée (ce n'est pas grave de faire des fautes); finalement, c'est l'intérêt de motiver les élèves à écrire qui l'a emporté.

Instructions

Temps

Ce cahier contient un exercice d'écrit. Pour l'ensemble de l'exercice, tu as **20 minutes** à ta disposition.

Dans ce laps de temps, tu dois:

- lire avec beaucoup d'attention la situation et l'exercice à faire;
- écrire un **texte** que tu pourras rendre **terminé** au bout de 20 minutes.

Remarque

Le plus important est qu'on **compre**ne ton texte.

N'aie donc pas trop peur de faire des fautes.

III. 15: Instructions générales pour l'exercice 906

2. Mise en situation de l'exercice, avec référence à une tâche réelle qui peut se présenter:

Des informations sur le chocolat

Situation

Tu as décidé de faire en cours d'allemand un petit exposé sur le chocolat. Tu aimerais donc demander des informations à la fameuse fabrique de chocolat suisse *Lindt*. Sur Internet, tu as trouvé une page qui propose un formulaire de contact. Tu le remplis et tu l'envoies à *Lindt*.

III. 16: mise en situation de l'exercice 906

3. Lignes directrices: grâce à ces lignes, les élèves n'ont pas vraiment eu besoin de se soucier de la structure conceptuelle du texte. Cette restriction se justifie par plusieurs raisons: le peu de temps dont disposaient les élèves n'aurait pas été suffisant, pour nombre d'entre eux, pour planifier la structure de la réponse, puis rédiger un texte qui aurait été le reflet de cette planification. Lorsque les contenus (et les fonctions) sont imposés, les travaux réalisés sont plus faciles à comparer. Il s'avère moins aisé d'avoir recours à des stratégies d'évitement, si bien que les déficits dans le domaine des ressources (compétences) langagières sont plus souvent observables. Comme désavantage possible, on pourrait citer le fait que la capacité d'un élève à créer, sur la base de sa propre planification, des textes qui sont adaptés à une mise en situation générale imposée peut moins bien s'exprimer⁶⁰. De plus, il faut s'attendre à ce que les élèves réalisent des textes qui, en quelque sorte, répondent aux lignes directrices

⁶⁰ Le peu de temps mis à disposition, la restriction imposée par le nombre relativement élevé de lignes directrices et, jusqu'à un certain point, l'interdiction d'avoir recours à des ouvrages de référence peuvent être vus comme une limitation de la validité cognitive (cf. Shaw / Weir 2007, 34 sqq.) des tests d'écriture, puisque les compétences de planification d'une part et, de l'autre, les compétences de remaniement ne peuvent se manifester que dans une certaine limite.

de manière séquentielle. Néanmoins, on a considéré que les avantages présumés l'emportaient sur les inconvénients potentiels.

Devoir

Ecris une courte demande à la fabrique de chocolat *Lindt*.

Dans ta demande, tu dois écrire quelque chose en allemand sur les points suivants:

- Pourquoi tu écris (ton exposé, ton intérêt, ...)
- Ce que tu aimerais savoir ou obtenir pour ton exposé sur le chocolat (cite au moins 3 points)
- Où l'entreprise *Lindt* peut t'envoyer du matériel (ton adresse postale)

Tu peux encore écrire autre chose, si cela va bien avec ton texte.

III. 17: Lignes directrices pour l'exercice 906

Comme le montre l'exemple, les exercices d'écriture ne sont pas présentés dans la langue cible, mais entièrement dans la langue scolaire locale, contrairement à ce qui est courant dans les examens internationaux connus ou à ce qui est habituel (jusqu'à présent) dans les examens allemands et autrichiens sur les standards de formation. On peut avancer différents arguments qui appuient cette décision: 1) dans des exercices de tests de ce genre, qui sont orientés sur l'action, la donnée joue un rôle important. Il ne s'agit pas simplement de rédiger un texte quelconque; il est question d'accomplir une tâche de communication réalisant des buts imposés. La mise en situation et la donnée remplacent ce qui, dans des occasions d'écriture réelles, va souvent de soi: le contexte opérationnel et l'intention de l'écriture. 2) Le but n'est pas de tester les compétences en écriture en fonction des compétences en compréhension écrite; à l'inverse, la mise en situation et la donnée ne doivent pas fournir des moyens linguistiques qui pourraient être repris pour l'écrit; cela d'autant plus que les élèves n'ont pas du tout les mêmes habitudes, suivant la pratique de leur enseignante ou enseignant, en ce qui concerne les moyens linguistiques à disposition (certains enseignants interdisent, d'autres encouragent leur utilisation). 3) Contrairement à une situation d'enseignement normale, les élèves doivent se débrouiller tout seuls dans les exercices des tests HarmoS. De plus, la compréhension de l'exercice doit prendre le moins de temps possible, étant donné le peu de temps mis à la disposition des élèves. Contrairement à HarmoS, les examens de diplômes internationaux peuvent en règle générale partir du principe que les candidats connaissent bien les formats de réponse possibles et le langage des instructions, puisqu'ils s'y sont spécifiquement préparés.

L'élaboration concrète des exercices s'est passée en plusieurs étapes. Une petite équipe (Ruth Keller et Peter Lenz) a rédigé, sur la base des *construct maps*, une série d'exercices qui ont été présentés en plusieurs cycles au consortium. Dans ce processus, les exercices ont été régulièrement modifiés; certains ont aussi été éliminés.

Enfin, des tests préliminaires ont été réalisés dans 18 classes d'allemand, de français et d'anglais. De plus, trois didacticiennes / didacticiens spécialisés dans les langues étrangères ont été consultés en Suisse romande. Les exercices ont alors encore pu être optimisés, sur la base des textes d'élèves ainsi rédigés, de l'analyse des questionnaires que les enseignantes et enseignants ainsi que les élèves des classes d'essai devaient remplir, et grâce aux suggestions soumises par les didacticiennes / didacticiens spécialisés.

Finalement, les exercices ont été traduits en français et ajoutés à la série des exercices de tests.

Il s'est avéré par la suite, durant l'enquête elle-même, que, malgré toutes les mesures prises, l'exercice 602 («Envoyer une carte d'anniversaire») était compris de façon tellement variée par les élèves de sixième année que les textes qui en ont résulté n'ont pas pu être mis à profit dans l'enquête HarmoS sur l'écrit.

3.6.3 Collecte des données

Dans le cadre de l'enquête de validation organisée par la CDIP, les élèves ont dû rédiger au maximum deux textes dans une langue étrangère, soit dans la même langue, soit dans deux langues différentes. Chaque exercice a été fait par 150 à 250 élèves, suivant la langue et le groupe dans lequel il a été proposé. Etant donné que les cahiers de tests ont été différenciés au niveau des élèves, on a obtenu un très grand nombre de combinaisons d'exercices dans chaque langue et dans les trois langues testées la même demi-journée (L1 et deux langues étrangères). La durée d'exécution s'élevait, pour huit des neuf exercices d'écriture dans les langues étrangères, à 20 minutes; pour l'exercice 912 («Interdiction des téléphones portables») à 30 minutes. Aucune aide n'était autorisée. En raison du dispositif choisi pour l'enquête (la responsabilité pour l'exécution des tests a été entièrement déléguée aux enseignantes et enseignants, et les consortiums n'ont eu aucun accès aux classes prenant part à l'enquête), on sait très peu de choses sur le déroulement de la matinée de test et sur les réactions des élèves ainsi que des enseignantes et enseignants face aux modalités des tests et aux données des exercices. Outre les productions des élèves, on dispose uniquement d'un petit nombre de données sociologiques, qui ont été prélevées au moyen d'un bref questionnaire distribué aux élèves.

3.6.4 Evaluation

A la suite de l'enquête principale, les textes ont dû être évalués selon certains critères. On opère alors une distinction entre l'**évaluation holistique** et l'**évaluation analytique**. Les deux évaluations ont été réparties dans le temps et réalisées indépendamment l'une de l'autre. Pour l'évaluation holistique des textes d'élèves, les évaluateurs avaient pour tâche de parvenir à une seule évaluation globale pour chaque texte, en tenant compte de certains points prédéfinis. Pour l'évaluation analytique, ces points et d'autres encore ont été analysés et concrétisés, puis saisis séparément par les évaluateurs.

Les ressources étant limitées, en temps et en personnel, il ne nous a pas été possible d'évaluer la totalité des textes. Lorsque, pour un même exercice⁶¹, nous avons reçu un nombre plus élevé que la moyenne de textes d'une même sous-population, nous avons éliminé, avant d'entamer l'évaluation holistique, un nombre relativement petit de textes sur la base d'un processus aléatoire. Après un premier examen, il s'est avéré que les travaux réalisés à partir de l'exercice 602 («Envoyer une carte d'anniversaire») ne pouvaient pas du tout être retenus. Quant aux travaux sur les quatre exercices 801 («Se présenter brièvement»), 807 («Un jour d'école ordinaire»), 808 («Expérience avec un animal domestique») et 811 («Avis sur les devoirs»), ils ont aussi été évalués par voie *analytique* dans toutes les langues testées, à l'aide d'un catalogue de critères assez long (voir ci-dessous). Précisons toutefois qu'il n'a pas été possible de procéder à une évaluation analytique lorsque le texte faisait moins de 25 mots environ. De ce fait, un nombre de textes plus ou moins important (selon les exercices) a dû être écarté, surtout parmi les textes des élèves de sixième année. Au final, le nombre de textes évalués par procédé analytique va de 100 à plus de 200 par exercice / langue cible. Ces textes proviennent, suivant l'exercice et la langue cible, de classes de 6^e et de 9^e année, ou parfois uniquement de l'un des deux niveaux.

L'évaluation holistique peut être considérée comme la colonne vertébrale de l'enquête. La base de données est structurée de manière à pouvoir représenter tous les élèves, à l'aide d'une analyse de *Rasch* (à plusieurs facettes), sur une seule échelle, «équitable», qui tient compte des différents degrés de difficulté des exercices et de la sévérité des différents évaluateurs. L'évaluation analytique visait à dégager des éléments assez différenciés, notamment dans le but d'optimiser les instruments d'évaluation pour de futures enquêtes; de plus, il s'agissait de créer une base de données dans le but d'identifier, dans le cadre d'études plus poussées, des schémas caractéristiques dans les profils de compétences des apprenants.

Lors de l'**évaluation holistique**, on a chaque fois eu recours à une échelle avec un barème allant de 0 à 5, la note «0» n'ayant été attribuée que lorsque rien n'avait été écrit, ou du moins rien qui

⁶¹ Le nombre inégal de textes d'élèves par exercice et par sous-population est essentiellement dû à la rotation des items: lorsque plus d'élèves devaient répondre à moins d'exercices dans le temps total imparti, on a obtenu davantage de travaux d'élèves sur le même exercice. Ainsi, les élèves de sixième année de Suisse romande ont reçu, dans l'ensemble, un spectre d'exercices moins large à accomplir durant la demi-journée de test en langues que les autres groupes ayant participé aux tests.

allait dans le sens de ce qui était demandé dans l'exercice. Si les notes 1 à 5 ont été fixées de manière spécifique à chaque exercice, elles ont toutefois été définies de manière aussi analogue que possible pour tous les exercices. Les différents points qui ont été inclus dans les descriptions associées aux notes concernent d'une part des aspects fonctionnels et pragmatiques et, d'autre part, des éléments linguistiques. La note 2 de l'échelle d'évaluation pour l'exercice 807 a par exemple été décrite ainsi:

2 Un texte recevant cette note présente une bonne partie des caractéristiques suivantes:

- texte court, parfois fragmentaire, qui produit un effet de communication limité
- description élémentaire et / ou en partie incohérente, dans laquelle quelques aspects sont mentionnés
- idées formulées simplement, suivant souvent la même structure de phrase, simple
- comporte un spectre limité de moyens linguistiques élémentaires, contient plusieurs fautes qui brouillent le message

Les évaluateurs ont été chargés de déterminer quelles étaient les descriptions associées aux notes (0, 1, 2 ...5) qui correspondaient *le mieux* à chaque texte évalué. Pour l'évaluation des textes dans les différentes langues cibles, on a utilisé les mêmes échelles pour un même exercice, de sorte que finalement, ce sont huit échelles holistiques qui se sont profilées, différant les unes des autres par quelques concrétisations spécifiques aux exercices.

Vu que les notes d'évaluation 1 à 5 sont décrites en fonction des exigences de chaque exercice, un élève donné devrait en principe obtenir une meilleure évaluation dans un exercice assez facile que dans un exercice plus difficile, pour autant qu'il ait réussi à mobiliser de la même manière ses compétences dans les deux exercices. Dans ce procédé, il ne s'agit donc pas d'un classement en fonction d'une échelle de référence commune pour tous les exercices, par ex. les niveaux du *Cadre européen de référence*. Ce rapport-là est établi dans un deuxième temps, sur la base de la mise en l'échelle, par voie empirique, de tous les textes réalisés à partir des différents exercices. La raison principale qui justifie ce procédé indirect vient du fait que, de cette manière, il est possible – vu qu'une partie des élèves a fait deux exercices d'écriture à chaque fois – de quantifier le facteur «difficulté des exercices», en ayant recours à une échelle de *Rasch*. Des évaluateurs humains ne seraient probablement pas en mesure d'estimer, avec la certitude souhaitée, l'influence de la difficulté des exercices sur les performances, ni d'en tenir suffisamment compte, comme d'un élément de relativisation, dans leurs évaluations. Pouvoir dégager des assertions, vérifiées empiriquement, sur la difficulté des exercices constitue une étape importante en vue de la validation du *construct* et un pas préalable important pour de futures spécifications d'exercices.

En général, pour l'évaluation holistique, quatre évaluateurs se tenaient à disposition par exercice pour chacune des trois langues cibles. Comme la codification a pris plus de temps que prévu initialement, les évaluateurs n'ont pas tous pu participer jusqu'à la fin. En revanche, quatre autres évaluateurs ont ensuite rejoint l'équipe. Pour chaque exercice, une partie des textes a chaque fois été évaluée, par langue cible et par sous-population (6^e ou 9^e années / CH alm. ou CH rom.), par tous les évaluateurs. On a visé une évaluation commune de 15 textes par combinaison langue cible / population, ce qui signifie notamment que, pour les exercices qui ont été faits tant dans des classes de sixième que dans des classes de neuvième année, on a pu parvenir à une trentaine de textes évalués par tous. L'évaluation multiple des performances par différents évaluateurs permet une estimation de la fiabilité inter-examineur, qui se compose à son tour de la cohérence (interne) de chaque évaluateur et de la sévérité dont il fait preuve dans son évaluation. Les analyses ont montré que les évaluateurs portent des jugements parfois assez différents quant au degré de sévérité. Néanmoins, vu que la sévérité des évaluateurs peut être incluse, comme une «facette» dans l'analyse statistique avec le logiciel *Rasch* «ConQuest», et qu'elle peut être largement neutralisée dans la mise en échelle, cela ne pose problème que dans les cas extrêmes (aux extrémités de l'échelle, lorsque par exemple 1 et 2 sont nivelés vers 1). Par contre, la cohérence de l'évaluateur doit être garantie dans le processus d'évaluation déjà, puisqu'aucune correction de cet aspect n'est possible ultérieurement.

Des mesures d'assurance qualité ont été prises dans ce sens durant le processus d'évaluation: une phase d'entraînement à l'évaluation («rater training») a eu lieu au début du processus, puis à chaque fois que les évaluateurs se mettaient à travailler sur de nouveaux exercices. Les principaux objectifs poursuivis dans ce contexte ont été de permettre aux évaluateurs d'adopter une compréhension des échelles qui corresponde à l'intention des auteurs, et de préparer une application sûre des critères pour les différents textes d'élèves. Dans un premier temps, les expériences des évaluateurs ont aussi chaque fois été mises à profit pour améliorer encore les échelles, et surtout pour lever les ambiguïtés éventuelles dans les formulations. Pour éviter que, après une concordance initiale, les interprétations des échelles ne divergent trop vite, on a repris, après une courte période d'évaluation individuelle, des textes à évaluer ensemble, puis une fois encore quand la moitié du travail environ avait été effectuée; les résultats ont ensuite été discutés au sein des groupes de travail spécifiques aux langues. La concordance entre les évaluateurs a aussi régulièrement fait l'objet de vérifications durant les premiers jours, au moyen de petits tests statistiques sur la fiabilité inter-examineur (*kappa* de Cohen pour deux évaluateurs à chaque fois). Lorsqu'on remarquait quelque chose chez un évaluateur, on abordait la question avec la personne concernée⁶².

Les mesures prises pour la coordination de l'évaluation entre les langues ont été moins nombreuses qu'au sein des langues elles-mêmes, notamment pour des raisons de faisabilité – mis à part un départ commun, tout au début du travail d'évaluation, puis à l'introduction d'un nouvel exercice (et d'une nouvelle échelle). Le principal lien entre les langues provient des échelles d'évaluation identiques par exercice et des débats des «groupes de langues» avec les mêmes responsables membres du consortium.

Les évaluateurs ont à chaque fois travaillé sur des textes dans une seule langue. Sur la base des observations faites durant les sessions d'évaluation, nous pensons pouvoir affirmer que les évaluateurs n'ont, en moyenne, pas fait preuve de la même sévérité suivant les langues. C'est pourquoi on ne peut pas simplement mettre au même rang la sévérité moyenne des évaluateurs par langue pour établir des moyens de comparaison entre les langues. Des possibilités de comparaison ne se font jour qu'avec le classement des trois échelles spécifiques aux langues, définies par voie empirique, en fonction de l'échelle de référence européenne (pour le procédé, voir plus bas).

L'évaluation analytique s'est déroulée de la même manière que l'évaluation holistique. Toutefois, comme nous l'avons déjà mentionné, il a fallu opérer un tri, en raison du manque de temps à disposition, entre les exercices et les textes reçus. De plus, il n'a parfois pas été possible d'obtenir des évaluations multiples de textes en quantité suffisante, c'est-à-dire permettant une exploitation systématique. En conséquence, les résultats de l'évaluation analytique ne se prêtent – comme c'était prévu dès le départ – qu'à des enquêtes complémentaires, principalement exploratoires, et non pas à des conclusions générales sur les populations d'élèves, par exemple.

Les différents items des instruments d'évaluation analytique se retrouvent dans les points suivants:

- 1) effet de communication (global) et pertinence du texte
- 2) adéquation et degré d'élaboration du texte à partir des points de contenu
- 3) aspects pragmatiques au niveau du texte:
 - longueur du texte
 - structure du texte
 - développement des thèmes
- 4) aspects lexicaux

⁶² Les interventions au cours d'une évaluation peuvent poser problème car elles comportent le risque que les évaluateurs se montrent tout à coup plus sévères ou moins sévères, en raison des comparaisons établies, dans les jugements qu'ils portent. Cette forme d'incohérence ne peut être saisie directement par des moyens statistiques. Le consortium a opté pour la forme d'accompagnement du processus d'évaluation présentée ci-dessus, estimant qu'il était plus important et plus efficace de chercher à empêcher l'apparition de points de vue divergents sur les textes des apprenants, plutôt que de saisir (seulement) certains effets après coup, par des méthodes statistiques.

- spectre lexical
- correction lexicale
- aspects linguistiques de la structure du texte (cohésion)

5) aspects grammaticaux

- domaine verbal
- domaine de la phrase

Les grilles d'évaluation relatives aux quatre exercices qui ont également fait l'objet d'une évaluation analytique sont identiques, dans la mesure du possible, ou du moins analogues. Outre les aspects ayant trait au langage, aux niveaux pragmatique et linguistique, le contenu est aussi saisi de manière détaillée (sur la base des lignes directrices définies dans les données d'exercices), et ce dans le but de faire quelque peu la lumière sur le rapport généralement assez «encombrant» entre le contenu et la qualité du texte. Les critères qui se rapportent à des caractéristiques linguistiques sont formulés à un niveau relativement concret et concernent par exemple l'emploi des temps, les formes de phrases utilisées ou les moyens de cohésion déployés. On souhaitait que par ex. les formulations parfois peu concrètes des critères contenus dans le *Cadre de référence* («structures grammaticales simples») soient saisies, sur la base des résultats obtenus, de manière plus explicite et puissent être proposées pour de futurs instruments. Elles devraient aussi pouvoir être employées pour identifier, dans le cadre d'enquêtes plus poussées, des profils de compétences, évtl. spécifiques à la langue cible ou à la langue source, chez les élèves.

3.6.5 Résultats de l'enquête

La mise en échelle des données a pris plus de temps que prévu initialement, tant pour les évaluations holistiques que pour les évaluations analytiques. C'est pourquoi il a fallu, en raison des contraintes temporelles qui accompagnaient le projet, inverser le procédé qui aurait semblé naturel: la sélection des textes d'élèves pour illustrer les propositions de standards de base n'a pu (dans un premier temps) se faire sur la base des échelles empiriques; ce tri a dû s'effectuer sur la base des évaluations des experts. A cet effet, ils se sont appuyés sur les descriptions de niveaux contenues dans le document relatif aux standards de base HarmoS, qui reposent sur différents textes voisins du *Cadre de référence*, ainsi que sur les textes *de référence* que le Conseil de l'Europe a publiés pour illustrer les niveaux de référence.

Ce n'est que par la suite que les trois échelles établies par voie empirique pour l'allemand, le français et l'anglais comme langues étrangères (à chaque fois une échelle distincte) ont été mises en regard des niveaux de référence européens. Les trois échelles se basent sur l'évaluation holistique. Pour le classement en fonction du *Cadre de référence*, on a à nouveau eu recours à un procédé qui repose sur le consensus des experts, mais qui prend comme point de départ, à l'instar de la procédure connue sous le nom de *bookmark standard setting*, l'échelle empirique de tous les textes dans une langue. Concrètement, les experts se sont fixé comme objectif d'identifier, sur l'échelle, au moins deux points éloignés l'un de l'autre, auxquels ont lieu des transitions de niveaux. A l'aide des exemples concrets, il s'est avéré qu'aux points de transition entre les niveaux A1.2 et A2.1 ainsi que B1.1 et B1.2, il y avait suffisamment de travaux d'élèves qui convenaient pour pouvoir fixer, avec une certaine assurance, les *bookmarks* (litt.: «marque-pages») en question. Ensuite, à partir de ces points fixes, les autres transitions de niveaux ont été déterminées de façon purement mathématique, en respectant des intervalles réguliers⁶³. Les attributions de

⁶³ Dans la méthode du *bookmark standard setting*, on ne se base pas généralement sur des distances égales entre les transitions de niveaux. Toutefois, ce procédé se justifie en raison des propriétés des deux échelles: l'échelle de référence européenne a été établie par une mise en échelle de *Rasch* et constitue de ce fait une échelle à intervalles linéaire. De plus, les niveaux comportent sur cette échelle chaque fois autant de *logits* (unités dans l'échelle de *Rasch*) dans le domaine considéré. L'échelle empirique pour l'écrit est une échelle du même type qui, du point de vue psychométrique, est certes seulement plus ou moins parallèle à l'échelle de référence et dont les unités ne peuvent être directement comparées, mais qui peut (et qui doit, si l'on veut que la mise en regard soit légitime) quand même être considérée comme suffisamment égale pour que les distances relatives entre les niveaux puissent être reprises.

niveaux qui en résultent pour les différents travaux ont ensuite été vérifiées, quant à leur plausibilité, au moyen d'échantillons. Grâce aux liens établis entre les trois échelles et l'échelle de référence (niveaux de référence européens), les résultats qui ont été obtenus dans les trois langues (étrangères) peuvent faire l'objet de comparaisons quantitatives directes. Toutefois, dans toute comparaison entre les langues, il faut garder à l'esprit l'imprécision de la procédure d'*equating* qui a été choisie.

Comme cela était déjà le cas pour la compréhension de l'écrit, les résultats sont exprimés sur une «échelle HarmoS», qui s'oriente selon la répartition empirique des résultats en allemand et en français (6^e et 9^e année)⁶⁴. Voici les correspondances établies entre cette échelle (ligne inférieure du tableau) et les niveaux de référence européens, de même que les niveaux HarmoS:

<i>Cadre européen de référence</i>	A1.1	A1.2	A2.1	A2.2	B1.1	B1.2	B2 et plus
HarmoS 8^e année	I₈	II₈	III₈	IV₈	V₈		
HarmoS 11^e année	I₁₁			II₁₁	III₁₁	IV₁₁	V₁₁
Echelle HarmoS pour l'écrit (500/100)	(477) – –	– 527 – –	– 577 – –	– 627 – –	– 677 – –	(727) – –	

Les cases surlignées en gris correspondent aux standards minimaux suggérés par les experts pour les niveaux de 8^e année (l'actuelle 6^e année) et de 11^e année (l'actuelle 9^e année). Les valeurs numériques relatives aux transitions de niveaux sont indiquées pour l'échelle HarmoS de l'écrit car elles sont significatives pour les standards minimaux: ce qui est attendu, c'est «un bon A1.2» pour les classes de 8^e année, et «un bon A2.1» pour les classes de 11^e année. Les valeurs moyennes pour les différents niveaux peuvent être aisément calculées en cas de besoin.

Comme dans la section sur la compréhension de l'écrit, quelques résultats sont présentés ci-après en rapport avec la distribution des compétences en écriture au sein des populations testées. Ces résultats sont aussi mis en regard des propositions des experts pour les standards de base, de façon à montrer dans quelle mesure ils seraient déjà atteints aujourd'hui, c'est-à-dire avant que les nouveaux scénarios pour l'enseignement des langues étrangères aient déployé leurs effets.

⁶⁴ Le résultat moyen obtenu en allemand et en français correspond à 500 points; l'écart-type de la distribution équivaut à 100 points.

Exercices d'expression écrite	Français
Suisse alémanique	
6^e année	
Standard de base 1^{re} / 2^e LÉ (A1.2)	527
6^e année tous	462
Aujourd'hui, le standard de base serait atteint par 20 % des élèves de sixième année.	
6 ^e année – filles	471
6 ^e année – garçons	452
9^e année	
Standard de base 1^{re} / 2^e LÉ (A2.1)	577
9^e année tous	553
Aujourd'hui, le standard de base serait atteint par 40 % des élèves de neuvième année.	
9 ^e année – filles	568
9 ^e année – garçons	533
9 ^e année – exigences élémentaires	485
9 ^e année – exigences élevées	553
9 ^e année – exigences supérieures	630

Exercices d'expression écrite	Allemand
Suisse romande	
6^e année	
Standard de base 1^{re} / 2^e LÉ (A1.2)	527
6^e année tous	453
Aujourd'hui, le standard de base serait atteint par 20 % des élèves de sixième année.	
6 ^e année – filles	466
6 ^e année – garçons	441
9^e année	
Standard de base 1^{re} / 2^e LÉ (A2.1)	577
9^e année tous	536
Aujourd'hui, le standard de base serait atteint par 35 % des élèves de neuvième année.	
9 ^e année – filles	548
9 ^e année – garçons	523
9 ^e année – exigences élémentaires	447
9 ^e année – sans séparation	525
9 ^e année – exigences élevées	543
9 ^e année – exigences supérieures	579

Exercices d'expression écrite	
Suisse alémanique et Suisse romande	Anglais
9^e année	
Standard de base 1^{re} / 2^e LÉ (A2.1)	577
9^e année tous	578
9 ^e année – filles	592
9 ^e année – garçons	562
CH além.: 9^e année tous	584
Aujourd'hui, le standard de base serait atteint par 50 % des élèves de neuvième année.	
CH além.: 9 ^e année – filles	598
CH além.: 9 ^e année – garçons	567
CH além.: 9 ^e année – exigences élémentaires	532
CH além.: 9 ^e année – exigences élevées	591
CH além.: 9 ^e année – exigences supérieures	644
CH rom.: 9^e année tous	558
Aujourd'hui, le standard de base serait atteint par 40 % des élèves de neuvième année.	
CH rom.: 9 ^e année – filles	568
CH rom.: 9 ^e année – garçons	546
CH rom.: 9 ^e année – exigences élémentaires	478
CH rom.: 9 ^e année – sans séparation	525
CH rom.: 9 ^e année – exigences élevées	553
CH rom.: 9 ^e année – exigences supérieures	597

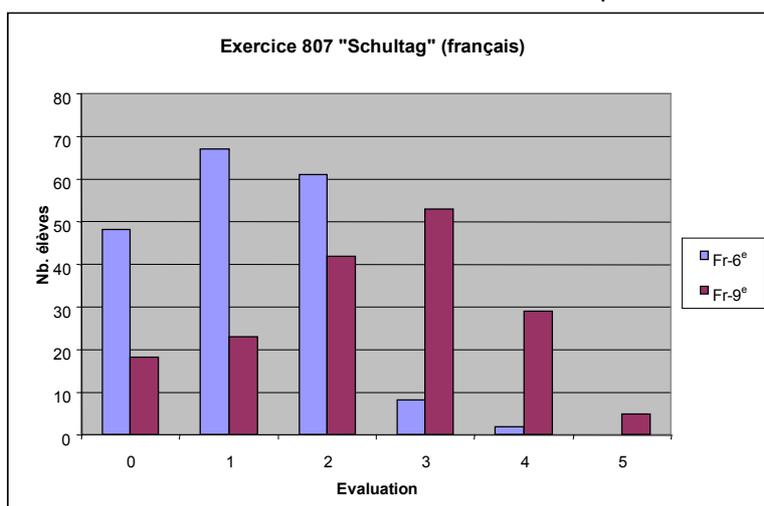
Comme nous l'avons déjà mentionné, le rapport entre les trois échelles repose sur une procédure d'*equating* réalisée grâce aux liens établis avec les niveaux de référence; nous avons par conséquent renoncé à proposer ici des comparaisons directes *entre* les langues (avec indication des significations). Les résultats numériques confirment toutefois les tendances qui se sont profilées au cours du processus d'évaluation, notamment l'observation selon laquelle les élèves de neuvième année ont rédigé de meilleurs textes en anglais que dans les langues nationales; c'est particulièrement le cas en Suisse alémanique, où l'on a récolté, dans l'ensemble, de meilleurs textes en anglais qu'en français. Dans la rédaction de textes en anglais, les jeunes Alémaniques présentent aussi des performances significativement meilleures que les jeunes Romands. Comme c'était déjà le cas pour la compréhension de l'écrit, les filles réalisent à nouveau des performances sensiblement meilleures, dans toutes les populations et sous-populations, que les garçons⁶⁵. Du reste, les résultats dans le domaine d'aptitudes de l'écriture peuvent, dans l'ensemble, assez facilement être comparés à ceux obtenus dans celui de la compréhension de l'écrit.

De manière générale, et notamment dans le cas des élèves de sixième année, on constate que les standards de base proposés seraient atteints actuellement par une minorité d'élèves uniquement,

⁶⁵ Il n'y a que dans les résultats en français pour la sixième année que l'écart observé n'est pas significatif. Dans les t-tests permettant de définir le niveau de signification, les données ont été pondérées sur la base des caractéristiques des échantillons d'élèves; on a également admis un effet de plan («design effect») de 2, qui s'est répercuté sous la forme d'une évaluation plus prudente des significations.

et ce malgré le fait que les standards pour la première langue étrangère en 8^e année et généralement en 11^e année sont fixés moins haut dans l'écrit que dans les autres aptitudes langagières de communication.

Ce résultat modeste en comparaison des standards proposés peut tout à fait être le reflet de compétences passablement faibles à l'écrit chez les élèves testés. Les résultats en moyenne très modestes s'expliquent certainement en partie, surtout pour les élèves de sixième année, par le fait que l'expression écrite avec intention de communication ne figure en général pas dans les objectifs d'apprentissage (prioritaires). Cependant, ce résultat est peut-être aussi en partie un effet des conditions difficiles dans lesquelles les élèves ont dû rédiger leurs textes. Le nombre de feuilles rendues vierges ou presque donne en effet à penser que, étant donné le peu de temps qui était accordé pour les exercices (20 minutes par texte), certains élèves se seraient simplement résignés, particulièrement parmi les plus jeunes. Les conditions difficiles dans lesquelles a eu lieu l'enquête pourraient donc avoir déclenché ce qu'on appelle un «effet plancher» (*floor effect*): des élèves qui, dans des circonstances favorables, auraient peut-être pu atteindre le niveau A1 ont pu être bloqués dans les conditions concrètes de l'enquête et n'ont pas pu montrer qu'ils auraient en fait pu réaliser de meilleures performances que d'autres. La distribution des points attribués initialement (à la place des mesures qui ont été exposées ci-dessus et qui, sur la base d'une mise en échelle, tiennent compte de la difficulté des exercices) nous permet de supposer qu'un tel effet s'est produit, vu la concentration extrême des résultats à gauche chez les élèves de sixième année (asymétrie positive [*positive skew*]; beaucoup de «0» et de «1» sur l'échelle allant de 0 à 5). Le graphique suivant montre avec quelle fréquence les différentes évaluations ont été atteintes par des élèves de sixième et de neuvième année pour un exercice plutôt simple.



III. 18: Distribution des évaluations par points pour des élèves de sixième et de neuvième année

De plus, on peut se demander si le classement des performances en fonction des niveaux de référence ne devrait pas, du fait des conditions difficiles rencontrées (pression par le temps inhabituelle, surtout pour les plus jeunes), être assorti d'un point d'interrogation: en effet, le lien entre les échelles empiriques et l'échelle de référence se base sur une comparaison des textes d'élèves (et de leurs caractéristiques) avec des descriptions de compétences en écriture et des paramètres textuels correspondant aux différents niveaux de référence. Ce rapport repose aussi sur une comparaison des textes d'élèves avec des textes de référence. Les conditions de réalisation des textes ne peuvent donc pas être prises en compte, alors même que l'expérience montre que des conditions difficiles peuvent avoir des effets systématiques sur les résultats. Cette éventualité devrait être considérée dans l'interprétation qui est faite des résultats HarmoS. Il n'est toutefois pas possible de procéder à une amélioration «technique» des résultats empiriques, puisque des données consolidées par rapport à l'ampleur de l'effet présumé font défaut.

Pour ce qui est des propositions de standards de base faites par les experts dans le domaine de l'écrit, on constate – comme d'ailleurs avec celles qui concernent la compréhension de l'écrit – qu'elles sont exigeantes et qu'elles ne peuvent assurément pas être atteintes par le nombre

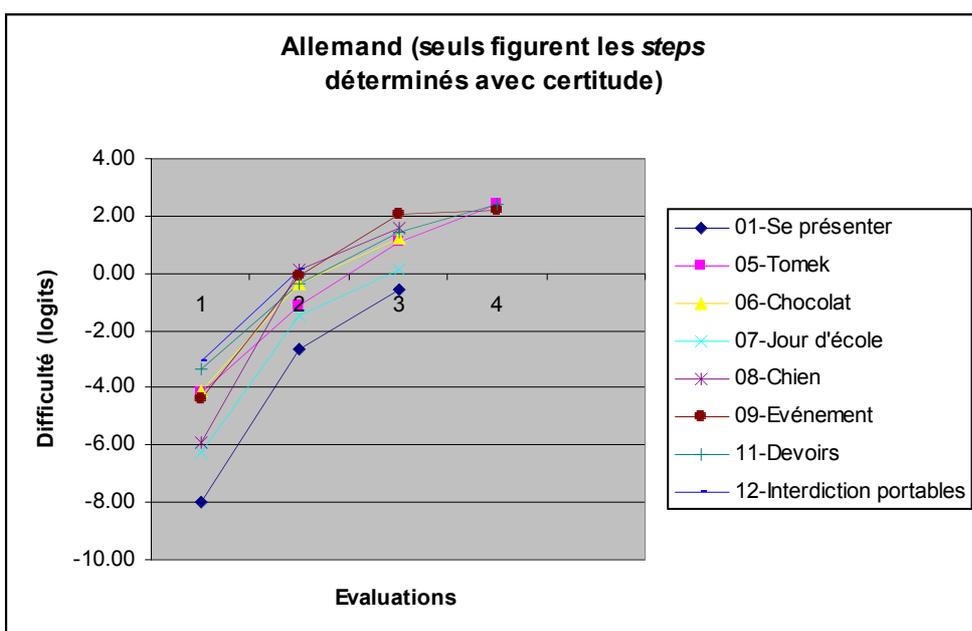
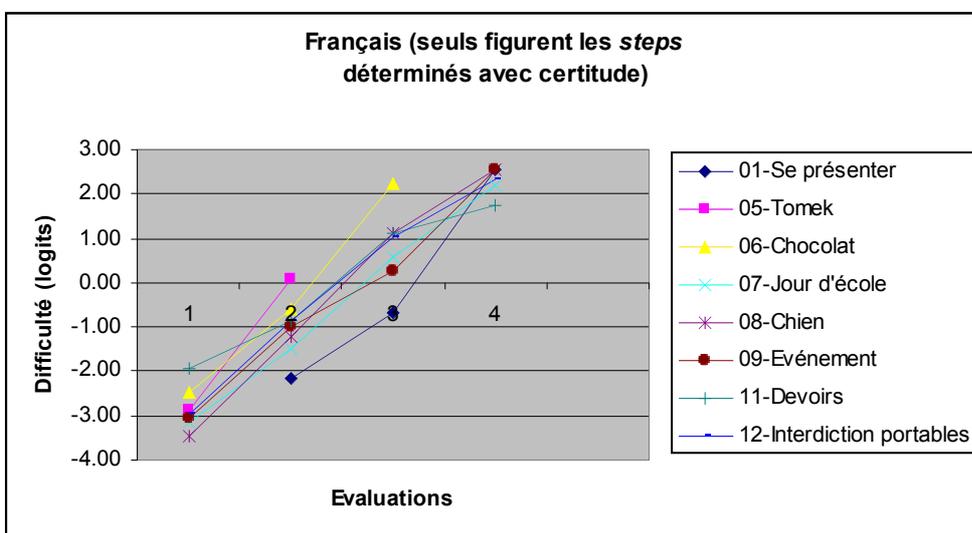
souhaité d'élèves si les nouveaux scénarios d'enseignement ne permettent pas d'obtenir des résultats sensiblement meilleurs qu'à l'heure actuelle.

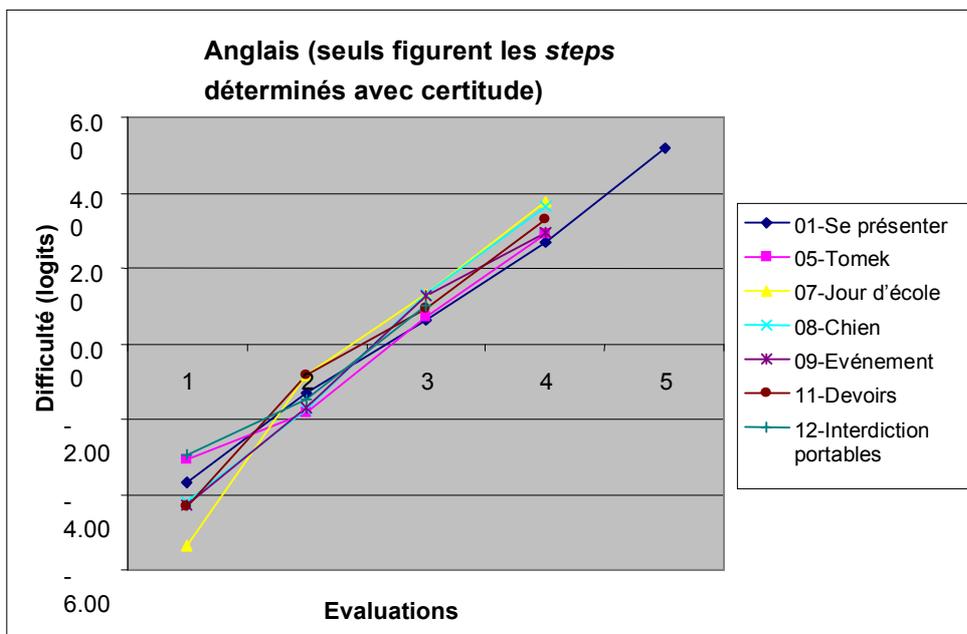
3.6.6 Analyses des résultats

Le *pool* de données d'HarmoS, vaste et diversifié, engage vraiment à effectuer bien plus d'analyses que ce qui a pu être réalisé jusqu'à présent. Nous nous contenterons ici d'aborder le facteur de la difficulté des exercices (a), ainsi que quelques conclusions tirées de l'analyse de l'évaluation analytique (b).

a) Difficulté des exercices

Comme nous l'avons déjà décrit, les tests d'écriture sur les langues nationales étaient composés de 9 exercices différents (dont 8 qui ont pu être exploités), et le test d'anglais de 7 exercices. On a estimé que les séries de tests permettraient, dans l'ensemble, d'obtenir un large spectre de compétences. L'évaluation des textes d'élèves au moyen de catégories holistiques, axées sur le produit et décrites en termes identiques ou analogues, a permis une évaluation de la difficulté des différents exercices grâce à une échelle de *Rasch* – ou plus précisément: une évaluation de la difficulté à atteindre les notes 1 à 5 dans les différents exercices. Les graphiques suivants présentent un aperçu des résultats:





Ill. 19: Comparaison des degrés de difficulté de tous les exercices dans les tests de français, d'allemand et d'anglais

A ce sujet tout d'abord un exemple de lecture et une remarque préliminaire «technique»:

Exemple de lecture: l'axe des ordonnées représente les unités de difficulté, ou de compétence, qu'on appelle les *logits*; sur l'axe des abscisses figurent les évaluations (théoriquement de 0 à 5), ou plus précisément les points auxquels une évaluation donnée devient plus probable que celle qui est immédiatement inférieure. Les informations sur l'exercice «07-Schultag» (français) peuvent se lire ainsi: pour qu'un élève obtienne, pour un texte relatif à cet exercice, la note «1» plutôt que «0», il doit disposer d'une compétence de -3.13 logits; pour avoir la note «2» plutôt que «1», sa compétence doit être de -1.49 logits (sur l'échelle de l'écrit en français – entre les langues, seuls les positionnements relatifs sont comparables avec cette forme de représentation). Avec une compétence égale, seul l'exercice «01-Se présenter» lui permettrait d'obtenir plus facilement un «2».

Remarque préliminaire technique: le manque d'indications sur la difficulté des différents *steps* (nœuds sur les lignes), plus élevés ou moins élevés, dans les différents exercices, s'explique par le fait que la catégorie de points immédiatement supérieure ou inférieure est trop rarement disponible pour pouvoir établir des échelles fiables⁶⁶. Le nombre d'évaluations à disposition a été sensiblement restreint par le fait que seuls les élèves qui avaient écrit deux textes dans la langue considérée pouvaient être utilisés de manière productive dans l'échelle pour dégager des indications sur la difficulté relative des exercices. Parmi les exemples notables d'exercices qui n'ont que peu de *steps* évalués statistiquement, citons, pour le français, l'exercice 05 «Tomek»⁶⁷; pour l'allemand (et, dans une moindre mesure, pour l'anglais) l'exercice 12 «Interdiction des téléphones portables». Les raisons pourraient être que l'exercice 05 («Tomek») en test de français n'a été proposé qu'à des élèves de sixième année, et que ceux-ci disposent d'un spectre de compétences limité. Quant à l'exercice 12 («Interdiction des téléphones portables»), l'élément déterminant pourrait être que l'exercice est assez difficile, ou du moins donne cette impression.

Pour ce qui est de la difficulté des exercices, on observe dans l'aperçu que, pour les langues nationales, où des élèves de sixième comme de neuvième année ont été testés, une hiérarchie de difficulté se profile, qui correspond dans l'ensemble à ce qui était attendu (assurément relativement simples: «01-Se présenter», «07-Un jour d'école ordinaire», «05-Tomek»; assurément plus difficiles: «11-Avis sur les devoirs», «12-Interdiction des téléphones portables»). Une exception se

⁶⁶ Les représentations choisies sont basées sur une mise en échelle des *item steps* (seuil de 50% entre deux catégories d'évaluation, par ex. «1» et «2»), avec le modèle *Rasch «partial credit»*.

⁶⁷ Dans cette section, ce sont les désignations de l'ill. 19 qui sont reprises. Contrairement à ce qu'on peut lire dans l'aperçu des exercices qui se trouve au début du chapitre 3.6.2, le premier chiffre (6, 8 ou 9) à chaque fois été omis.

fait jour dans les tests de français avec l'exercice «05-Tomek», la réponse attendue étant une lettre sous la forme d'un e-mail. Contrairement aux deux autres langues, cet exercice n'a été proposé, en français, qu'à des élèves de sixième année. Ces derniers ont à l'évidence (et à notre surprise) eu de la peine à faire cet exercice; ils ont même eu de la peine à atteindre la note «3». En outre, plus de 25% de ces élèves de sixième ont rendu une feuille plus ou moins vierge, obtenant ainsi la note «0». Comme cet exercice n'a pas été fait par des élèves de neuvième année, il manque en l'occurrence le «regard d'en haut» qui atténue souvent de tels effets (cf. à ce sujet le comportement de «05-Tomek» dans le test d'allemand, dans lequel près de 20% des élèves de sixième année ont obtenu la note «0», mais dans lequel les élèves de neuvième année ont visiblement rétabli les relations de difficulté). Dans le test d'allemand, les deux exercices qui n'ont été proposés qu'à des élèves de sixième année («01-Se présenter» et «07-Un jour d'école ordinaire») se sont comportés conformément aux attentes et paraissent être plus faciles que les autres exercices. Il est intéressant de noter que, dans le test d'anglais, auquel seuls des élèves de neuvième année ont pris part, les degrés de difficulté des exercices sont très proches. Du reste, dans l'exercice «01-Se présenter», la note «5» a été attribuée si souvent que c'est le seul exercice où cette note figure sur l'échelle. Par contre, dans l'exercice «12-Interdiction des téléphones portables», la note «4» n'a pas pu être reprise sur l'échelle par manque de données. Les faibles effets liés aux exercices dans le test d'anglais et le résultat surprenant de l'exercice «05-Tomek» en français pourraient indiquer que les exercices doivent être sélectionnés et formulés très soigneusement, surtout pour les élèves de sixième année, si l'on veut éviter qu'un «effet plancher» se produise pour des raisons qui ne relèvent pas de la substance même du *construct* pour l'écrit. A l'inverse, les résultats montrent également que les exercices qui visent une portée et un caractère opérationnels pour les niveaux inférieurs permettent encore des différenciations aux niveaux supérieurs. Cela s'explique probablement en partie par le fait que, dans les textes des exercices notés par exemple de quatre points, on doit trouver, au niveau linguistique également, plusieurs éléments qui pointent vers une évidence positive et, en même temps, peu d'éléments tendant vers une évidence négative. Les ressources qui doivent être mobilisées pour une certaine qualité et un certain effet du texte semblent largement se correspondre, indépendamment de l'exercice⁶⁸. Dans ce contexte, il n'est pas aisé d'interpréter la faible augmentation des valeurs de difficulté à l'extrémité supérieure de l'échelle en allemand. La progression parabolique (plutôt que linéaire) des courbes de difficulté et particulièrement les grands sauts que l'on constate entre les évaluations «1» et «2» semblent surtout indiquer une interprétation plus stricte des évaluations «2» et «3» par les évaluateurs de l'allemand, si bien que, en fin de compte, la différence de contenu entre les catégories «3» et «4» n'était plus que minime et que les deux notes ont été attribuées presque indifféremment.

b) Les évaluations analytiques n'ont, pour l'instant, été analysées de plus près que pour l'anglais, et ce surtout par rapport à la dimensionnalité qui apparaît dans les items analytiques. Les évaluations holistiques ont aussi été en partie reprises dans ces analyses, dans le but de repérer où se situent les évaluations holistiques en rapport avec les sous-dimensions (par ex. «contenu», «langue») qui se profilent potentiellement dans les items analytiques⁶⁹. Les analyses exploratoires avaient pour but de vérifier les instruments d'évaluation et de permettre une meilleure compréhension des évaluations. Il s'agissait aussi de dégager des indications, en vue des travaux qui seront entrepris à l'avenir, permettant une optimisation des instruments d'évaluation.

Concrètement, les évaluations ont été analysées à chaque fois séparément pour trois exercices d'anglais. Pour les analyses, on a travaillé – par une approche imbriquée – avec la méthode de l'analyse factorielle et l'échelle de *Rasch*, notamment des analyses de *Rasch* pluridimensionnelles du programme «ConQuest». Les trois exercices d'anglais ont livré des résultats assez semblables, au point qu'il est possible de faire certaines généralisations (du moins pour l'anglais). Voici un aperçu de quelques-uns des résultats qui ont été obtenus:

⁶⁸ C'est du moins ce qui semble se confirmer dans le cas de ces exercices et de l'échelle d'évaluation développée dans ce cadre. Cela ne nous permet toutefois pas de savoir dans quelle mesure ces conclusions peuvent aussi être reportées sur des exercices et des textes qui nécessitent une *CALP* (*communicative academic language proficiency*) bien développée.

⁶⁹ Les domaines qui devraient être saisis avec l'évaluation analytique sont mentionnés plus haut, dans la section 3.6.4.

- Une comparaison entre les évaluations holistiques et la totalité des évaluations analytiques pour les trois exercices révèle des corrélations (sans correction des erreurs de mesure) de 0.71, 0.74 et 0.83. Celles-ci sont légèrement supérieures lorsqu'on n'inclut pas tous les items analytiques dans la comparaison, mais qu'on prend seulement les trois items qui s'évaluent rapidement et facilement, à savoir 110 «effet de communication», 710 «spectre lexical» et 300 «longueur du texte»: 0.72, 0.75, 0.85. (Une comparaison entre ces trois items et la vingtaine d'items de l'évaluation analytique totale montre pour sa part une corrélation haute, d'au moins 0.91.) La longueur du texte seule (nombre effectif de mots exprimé dans les catégories d'intervalles 0 à 5) est en corrélation, avec quelques points de pourcentage de moins que l'ensemble des items analytiques, avec les évaluations holistiques. Des analyses factorielles relatives aux trois exercices étudiés, pour lesquels on a rassemblé les données holistiques et analytiques, montrent que les items 110, 300 et 710 suivent effectivement une orientation très semblable à l'évaluation holistique, et qu'ils font de plus très fortement appel à la dimension principale de *Rasch*⁷⁰.
- Tant les analyses factorielles que les calculs qui s'y réfèrent, réalisés à l'aide du logiciel «ConQuest» qui se prête particulièrement bien au travail avec plusieurs dimensions psychométriques, révèlent l'existence d'une certaine dimensionnalité au sein des items de l'instrument d'évaluation analytique, c'est-à-dire que l'on parvient à mesurer différentes dimensions de compétences. Ces dimensions peuvent être schématiquement caractérisées ainsi: contenu+structure du texte+longueur du texte vs. langue (vocabulaire et grammaire, y compr. cohésion). Les groupes d'items ainsi définis présentent, selon l'exercice, des corrélations de 0.78 à 0.85. Vu que «ConQuest» mesure des dimensions qu'on appelle «latentes», avec correction des erreurs de mesure (elles indiquent de ce fait la limite supérieure de l'intervalle d'estimation), on peut réellement parler de dimensions distinctes. Néanmoins, elles n'apparaissent pas assez distinctement pour qu'une échelle de *Rasch* unidimensionnelle, avec prise en compte des items qui saisissent ces dimensions probables, ne soit plus légitime. Au contraire, la dimensionnalité qui se profile devrait être une indication du fait qu'un instrument d'évaluation comporte effectivement des items des deux groupes et qu'ils devraient évtl. être pondérés de manière ciblée.
- Des analyses relatives au fonctionnement différentiel des items («differential item functioning» - DIF) ont également été réalisées parallèlement à la mise en échelle des items analytiques. On a constaté que le DIF n'est apparemment pas uniforme, suivant les éléments de compétences auxquels se rapportent les items: pour les élèves plus faibles, les items relatifs au contenu seraient en fait plus faciles, et les items linguistiques plus difficiles qu'ils n'apparaissent sur la base de l'échelle qui inclut toute la population des élèves. Cela va (probablement) de pair avec le constat (fait au cours de l'analyse de *Rasch*, où les dimensions étaient séparées) que les items relatifs au contenu causent moins de dispersion que les autres. Ces premiers résultats concernant ce DIF non uniforme devraient être approfondis, parce que ce phénomène pourrait avoir des conséquences sur la formulation des grilles de compétences, qui comportent tant des éléments de contenu que des aspects linguistiques.
- Pour l'évaluation analytique, on a formulé des items qui devaient permettre de saisir la compétence grammaticale de manière plus concrète que ce n'est fait par exemple dans la grille (applicable à plusieurs langues) utilisée pour évaluer la capacité de communication orale dans le *Cadre de référence*; il s'agit par exemple d'items qui se rapportent aux formes verbales ou aux types de phrases. Au cours de la mise en échelle (uniquement pour l'anglais), certains de ces items se sont révélés problématiques dans la mesure où ils présentaient un *fit* relativement insatisfaisant ou qu'on ne pouvait guère les classer sur une échelle parce que les phénomènes auxquels ils se référaient n'ont pu être observés que rarement (par ex. les négations). D'autres items, par exemple ceux relatifs aux phrases principales et subordonnées, se sont avérés productifs. Il serait intéressant de procéder à une comparaison transversale du fonctionnement des critères dans les trois langues étrangères qui ont été testées.

⁷⁰ Les analyses ont été effectuées surtout à l'aide de *Winsteps*. Outre les résultats des analyses en composantes principales (*principal components analysis*), les faibles valeurs *infit* donnent aussi à penser que ces quelques items reflètent une bonne partie de ce qui est saisi dans l'ensemble par l'évaluation holistique et analytique fusionnée.

3.6.7 *Etat des travaux et perspectives*

Comme il ressort implicitement et explicitement des résultats présentés ci-dessus, les données provenant de l'évaluation analytique fournissent une bonne base pour des enquêtes supplémentaires plus poussées, surtout du fait que les données disponibles proviennent de trois tests, qui portent en plus sur des langues différentes. En effet, il est ainsi bien moins probable de tirer des conclusions hâtives, basées sur des faits observés qui seraient dus au hasard, que si l'on ne se réfère qu'à une seule base de données.

Parmi ces enquêtes possibles, nous estimons que les travaux suivants notamment pourraient valoir la peine d'être menés et s'avérer importants:

- la finalisation d'un instrument d'évaluation qui permette de saisir de manière efficiente, sur une échelle complète, les aspects essentiels au point de voir apparaître un jugement d'ensemble correct.

A cet effet, il s'agit notamment d'identifier d'autres éléments essentiels pour la description des aspects grammaticaux des performances d'élèves dans différentes langues et à divers niveaux, ainsi que de décider de la pondération des différents éléments de compétences.

- Une correction et un accroissement du *pool* d'exercices: il devrait y avoir, à l'intention des apprenants plus jeunes et plus faibles, plus d'exercices qui leur permettent de montrer ce dont ils sont capables dans le domaine de l'écrit (qui s'effectue en fin de compte à des fins) de communication.

Il faut aussi se demander s'il est vraiment judicieux d'utiliser autant d'exercices (7 resp. 9), dans le but de couvrir le *construct* et spécifiquement de tenir compte de différents types de discours. Il serait sûrement difficile de trancher cette question sur la base des résultats empiriques disponibles, parce que, avec environ deux exercices par type de discours, on ne dispose pas d'une base suffisante pour tirer des conclusions permettant des généralisations. Les différences constatées entre les résultats des différents exercices ne peuvent probablement pas être attribuées avec certitude aux types de discours respectifs; elles peuvent tout aussi bien être liées au thème ou à la complexité du contexte.

- Des assertions sur les profils de compétences de certains sous-groupes des populations testées. Les évaluations analytiques offrent la base de données nécessaire pour dresser une description, dans le sens d'une différenciation, de profils de compétences de différents types d'apprenants sur un domaine de niveaux relativement étendu⁷¹.

3.7 *Parler*

3.7.1 *Situation initiale et bases de travail*

Le domaine d'aptitudes «Parler» constitue déjà actuellement l'une des priorités déclarées de l'enseignement des langues étrangères à l'école – au degré primaire encore plus qu'au degré secondaire. Néanmoins, jusque-là rien n'est dit sur la façon dont l'expression orale est effectivement considérée, enseignée et évaluée. On ne sait en particulier pas dans quelle mesure la classe «agit» effectivement dans la langue étrangère durant l'enseignement, c'est-à-dire dans quelle mesure on fait plus que des exercices et des jeux de rôle dans la partie orale de l'enseignement. Et l'enseignement de l'oral ne perdra pas de son importance dans les nouveaux scénarios d'enseignement. L'expression orale joue un rôle central, tant dans le cadre d'une pédagogie des échanges que dans un enseignement de plus en plus axé sur des contenus conceptuels – des domaines qui constituent tous deux des tendances d'évolution possibles. Il se pourrait même que la production orale (*S'exprimer oralement en continu*) gagne en importance avec cette approche des contenus conceptuels.

Dans la perspective de ces scénarios relatifs à l'utilisation de la langue en rapport avec l'enseignement, on a testé, dans l'expression orale comme dans l'écrit, la capacité d'action

⁷¹ Se basant sur les résultats HarMoS en anglais, un membre du consortium aborde, dans son travail de thèse, la question des différents (types de) profils de compétences d'une part et, de l'autre, celle de la description des performances dans le cadre d'une approche dite «novice-expert» (cf. Bereiter / Scardamalia 1987).

langagière et communicationnelle, à l'aide d'exercices (pour la majeure partie) clairement délimités sur le plan fonctionnel. Tous les exercices proposés nécessitaient une expression relativement spontanée, avec un accent sur le contenu⁷².

Le savoir et le savoir-faire pragmatique et stratégique revêtent une importance considérable dans un modèle sociocognitif, orienté sur l'action (recours à la langue dans différentes situations, adaptation de l'usage de la langue à la situation qui se présente; compensation, clarification, coopération, autocorrection – McKay 2006; Conseil de l'Europe 2000, Bachmann / Palmer 1996). Mais en même temps, l'aptitude d'expression orale comporte aussi, comme *construct* théorique, le savoir grammatical et textuel (connaissances en vocabulaire, syntaxe et phonologie; élocution cohérente et structurée), comme cela a par exemple été mis en exergue, de manière déséquilibrée, dans les jeux de rôle traditionnels, que l'on apprend par cœur.

Les tests HarMoS relatifs à l'interaction orale comprenaient différentes situations impliquant plus ou moins de personnes: d'une part, l'interaction avec des élèves du même âge que l'élève testé; d'autre part, une conversation avec un locuteur adulte, compétent, de la langue cible. Le but de ce procédé était de faire naître, du fait de ces différentes «combinaisons», d'autres interprétations de rôles et donc différentes nécessités, surtout par rapport aux compétences pragmatiques et stratégiques. D'après certaines études réalisées en la matière, l'aptitude cognitive et sociale qui permet de réaliser des interactions de manière contrôlée et de réagir à un vis-à-vis devrait, en principe, être pleinement développée chez un enfant de dix ans (McKay 2006, Cameron 2001).

Contrairement à ce qui a été fait pour l'écrit, l'interaction et la production ont été saisies séparément, parce que cette séparation est très ancrée dans la pratique et surtout parce que l'on pouvait supposer que les résultats ne correspondraient pas chez tous les élèves et que des différences apparaîtraient éventuellement en fonction du niveau de compétence (ou d'autres variables, telles que le développement cognitif ou l'âge). Contrairement à l'interaction, où le partenaire de discussion contribue en règle générale au processus de conversation, la production orale nécessite des aptitudes cognitives et linguistiques suffisamment développées pour que l'élève puisse maintenir une conversation de manière autonome, sur une certaine durée. Chez les jeunes apprenants, ces aptitudes dépendent à l'évidence fortement des compétences en littératie (*literacy skills*) qu'ils ont acquises dans la langue première (McKay 2006).

Comme cela a été décrit au chap. 3.2, 174 élèves en tout ont été testés sur leur expression orale, dans le but d'obtenir des points de repère relatifs au spectre de performances observable actuellement. Tous les élèves ont été testés, à l'aide des trois «combinaisons» de personnes déjà évoquées, quant à leurs compétences dans les deux domaines de la production orale et de l'interaction orale:

- Dans la production orale: monologue autonome;
- Dans l'interaction orale:
 - conversation avec un partenaire de discussion du même âge, en général d'une classe parallèle (*non-native-speaker peer interaction*);
 - entretien avec un locuteur adulte dont la L1 est la langue cible (*native-speaker / non-native-speaker interaction*).

En plus de cela, le test débutait par une prise de contact à trois locuteurs.

Dans la partie «production orale», les élèves devaient montrer avec quelle aisance ils étaient capables de produire une contribution orale d'une certaine longueur, basée sur une donnée écrite concrète et préparée en peu de temps seulement.

Dans la partie «interaction entre pairs» (*peer interaction*), les élèves devaient exécuter, à deux, un exercice d'interaction écrit, imposé en termes concrets, qui, suivant l'exercice, mettait en évidence soit des éléments de coopération, soit des éléments de confrontation.

⁷² Les productions orales non (ou peu) préparées se constituent par unités d'idées, et sont habituellement caractérisées par des phrases incomplètes, des mots de remplissage (*non-specific words*), qui ne contiennent pas ou que peu d'informations, ainsi que des pauses dans le discours. Les unités d'idées s'orientent à la capacité de mémorisation et de production du locuteur de même que, dans l'interaction, à celle de l'auditeur (Luoma 2004).

La conversation finale entre un élève et un locuteur natif (*native speaker*, appelé ci-après: interlocuteur) prenait essentiellement la forme d'un entretien. Contrairement à ce qui se passait dans les deux autres parties du test, les apprenants pouvaient alors compter sur les impulsions dans la conversation et le soutien offerts par l'adulte, plus compétent sur le plan linguistique. L'adulte avait pour tâche de sonder les limites linguistiques des élèves en adoptant un comportement adaptatif durant l'entretien.

3.7.2 Instruments (exercices)

Pour l'élaboration des exercices d'expression orale, on a essentiellement suivi les mêmes étapes et adopté les mêmes sources (documents et instruments relatifs au *Cadre de référence; lingualevel*) que pour les exercices d'écriture. A cela se sont ajoutées des contributions de la littérature spécialisée sur la compétence orale, ainsi que sur les tests et l'évaluation de l'expression orale⁷³.

On a à nouveau conçu des **construct maps**, dans lesquelles on a intégré des paramètres concernant les exercices et les performances pour différents niveaux de compétence. Elles ont servi de base pour l'élaboration des exercices et des échelles d'évaluation. Dans l'interaction et la production orales, on a également veillé à couvrir les différents types de discours dans des parties d'exercices adaptées en conséquence. On a toutefois fixé certaines priorités entre les différentes parties des tests: le type de discours «raconter / rapporter» n'a pas été mis en œuvre dans des exercices d'interaction orale entre élèves du même âge, parce que s'exprimer en continu, pendant un certain temps, constitue la priorité de la partie «production orale». Les types de discours «Maintenir la relation» et «Inviter / motiver à faire quelque chose» ont été principalement repris dans l'interaction entre élèves du même âge, dans le sens où ceux-ci devaient maintenir une conversation tout seuls et négocier quelque chose au travers d'un échange oral (par ex. se mettre d'accord sur un endroit idéal pour un camp scolaire, ou organiser une fête), ou encore demander à l'autre de faire quelque chose (par ex. passer un objet).

Le tableau suivant donne un aperçu des exercices, des types de discours visés et de la répartition des exercices en fonction des niveaux des classes.

⁷³ Cf. notamment Fulcher (2003), Kurtz (2001), Luoma (2004), McKay (2006), Weir (2005).

Production orale

	P1 Se présenter brièvement	P2 Portrait de classe	P3 Raconter une journée	P4 «Ta» ville	P5 TV – ordinateur – téléphone?	P6 La vie d'autrefois et d'aujourd'hui
Types de discours						
Informer / décrire	X	X	X	X	X	X
Inviter / motiver à faire quelque chose						
Raconter / rapporter			X	X	X	
Exprimer son opinion / argumenter				X	X	X
Maintenir la relation						
Exercices proposés en:						
6 ^e année (D / F)						
9 ^e année (D / F)						
Anglais, 9 ^e année, CH além. et CH rom.						

Interaction orale avec des élèves du même âge (*peer interaction, non-native*)

	IP1 Choisir et commenter des images	IP2 Camp scolaire	IP3 Préparer une fête d'adieu	IP4 Organiser des activités avec d'autres	IP5 L'uniforme à l'école?	IP6 Organiser une fête de classe
Types de discours						
Informer / Décrire	X	X	X	X	X	
Inviter / motiver à faire quelque chose	X		X			X
Raconter / Rapporter						
Exprimer son opinion / argumenter	X	X	X	X	X	X
Maintenir la relation						
Exercices proposés en:						
6 ^e année (D / F)						
9 ^e année (D / F)						
Anglais, 9 ^e année, CH além. et CH rom.						

Interaction orale avec des adultes (*native speakers*)

	Domaine thématique A famille / animaux domestiques météo / saisons loisirs vacances / voyages	Domaine thématique B télévision ou lecture? <i>fast-food</i> ou bio? santé technologie	Domaine thématique C rêves avenir gagner à la loterie! surprise
Types de discours			
Informier / Décrire	x	x	x
Inviter / motiver à faire quelque chose			
Raconter / Rapporter	x	x	x
Exprimer son opinion / argumenter		x	x
Maintenir la relation			
Exercices proposés en:			
6 ^e année (D / F)	1)	2)	2)
9 ^e année (D / F)	1)	2)	2)
Anglais, 9 ^e année, CH além. et CH rom.	1)	2)	2)

1) désigne l'ensemble des thèmes de départ pour tous les niveaux scolaires, 2) domaines thématiques plus difficiles. Comme l'entretien constituait une forme d'évaluation adaptative, lorsque les thèmes du domaine A avaient été traités, certains thèmes plus difficiles des domaines B voire C – niveau encore plus élevé – étaient abordés, en fonction de la performance orale de l'élève, indépendamment de son niveau scolaire.

La structure des exercices est comparable à celle des exercices d'écriture. Voici l'articulation suivie dans les exercices:

1) Mise en situation: l'élève doit pouvoir facilement se placer en pensée dans le contexte d'expression proposé:

Situation
On discute dans votre école pour savoir s'il faut introduire un uniforme scolaire. Votre avis nous intéresse!

III. 20: Mise en situation de l'exercice IP5

2) L'exercice lui-même, avec l'objectif d'expression visé, les lignes directrices sur les contenus attendus ainsi que des indications sur les types de discours imposés:

Devoir

1) **Décris** les vêtements que tu portes d'habitude pour aller à l'école.

2) **Discutez**: les uniformes scolaire **oui – non** ? **Avantages** et **inconvénients** ?

3) Choisis l'uniforme qui te plaît le plus (1 ou 2).
Convincs ton vis-à-vis des qualités de cet uniforme.

uniforme scolaire 1



uniforme scolaire 2



4) **Mettez-vous d'accord** sur **un** uniforme possible pour votre école.

III. 21: Donnée d'exercice structurante pour l'exercice IP5

Au travers de la prise de contact à trois et de l'interaction orale avec un locuteur adulte, il s'agissait de créer une situation d'expression spontanée aussi naturelle que possible, raison pour laquelle les élèves ne disposaient ni d'une feuille d'exercice, ni de temps de préparation. Durant la prise de contact, il y avait à chaque fois deux élèves. L'interlocuteur adulte leur a posé des questions

concrètes, dont l'ordre n'était pas prévisible pour les élèves mais correspondait à un principe précis (même nombre de premières questions et de questions subséquentes).

Pour l'interaction orale, chaque élève était seul avec un interlocuteur, qui était tenu de s'engager dans la discussion en montrant de l'intérêt et de s'adapter, dans le choix des moyens linguistiques et du tempo, au niveau de l'élève. Au départ, l'élève a pu choisir une carte parmi les quatre qui lui étaient présentées (domaine thématique A). Pour pouvoir sonder les limites de la capacité de performance langagière de l'élève, l'interlocuteur a fait augmenter, au cours de la conversation, le degré de difficulté à deux niveaux: d'abord au niveau des *prompts* et des questions (au départ, questions ouvertes, puis questions plus dirigées), ensuite au niveau des thèmes abordés (introduction avec le cercle thématique A: «informer et décrire» un sujet relevant de la sphère personnelle; cercle thématique B: exprimer son avis sur des thèmes (encore) familiers à l'élève; cercle thématique C: décrire des rêves et des éventualités dans le domaine de thèmes abstraits). Ce dispositif de test adaptatif a été conçu pour permettre à l'élève de montrer quelle est sa compétence langagière, sans aucune limitation qui serait imposée par une donnée d'exercice orientant fortement la structure de la réponse.

Les **exercices de tests** ont été développés par une petite équipe (Sandra Hutterli et Peter Lenz), sur la base de la *construct map*, et présentés, en plusieurs étapes, à un cercle plus large d'expertes et d'experts du consortium et du conseil consultatif; ils ont alors été soit adaptés en conséquence, soit écartés. Les exercices ont été testés dans la pratique, au cours d'une phase de *test préliminaire*, à l'aide d'un échantillon d'élèves provenant de deux classes différentes (classe primaire de 6^e année et classe secondaire de 9^e année), et les performances des élèves ont été filmées. Une fois les exercices terminés, les élèves ont répondu à des questions sur le déroulement des tests et donné leur avis personnel sur les exercices. L'analyse des réponses reçues et des enregistrements vidéo a fourni diverses indications sur les exercices:

- 1) Il est judicieux de garder la langue scolaire pour les instructions sur les exercices, puisque le temps de préparation est très court et que l'exercice doit être compris rapidement.
- 2) La mise en situation de l'acte de parole aide les élèves à s'imaginer être dans la situation. Elle doit toutefois être rédigée en quelques mots (entre une et trois phrases au maximum), et doit être directement orientée sur l'objectif visé.
- 3) Les éléments de contenu imposés sont compris le plus facilement lorsque les notions clés sont surlignées en couleur et que les aspects centraux sont illustrés par des images. Ces éléments contribuent à une élocution aussi libre que possible, détachée de la feuille d'exercice.
- 4) Les élèves ont réagi positivement à cette forme de tests et l'ont décrite comme leur donnant une possibilité de montrer leurs capacités. Parmi les points positifs, ils ont notamment relevé que les thèmes proposés leur laissaient une certaine marge de manœuvre, tant sur le plan de la langue que du contenu, et qu'on ne leur demandait pas des connaissances explicites.

Les adultes qui ont joué le rôle des **interlocuteurs** ont été choisis parce qu'ils étaient des locuteurs natifs (*native speaker*), qu'ils avaient le contact facile avec les enfants et les jeunes et qu'ils avaient eu l'occasion de travailler avec des personnes de langue étrangère. Mais, dans l'idéal, nous voulions éviter de prendre des personnes actives dans l'enseignement. Ce qu'on attendait d'eux, c'était un comportement sensible et encourageant; nous ne souhaitons pas une attitude scolaire, prompte à corriger les erreurs, ni trop fortement impliquée, surtout pas pour la dernière partie du test. De plus, il paraissait important que les interlocuteurs ne soient pas trop proches en âge des élèves de neuvième année, de façon à ce que les rôles soient intuitivement clairs. Ce dernier critère n'a toutefois pas été pleinement rempli par les interlocutrices de langue maternelle allemande; âgées de 25 ans, elles se sont parfois trop adaptées aux jeunes dans les entretiens menés. Un autre fait s'est aussi révélé problématique: par manque d'expérience avec le groupe cible, elles ont parfois adopté un allemand (trop) fortement influencé par des dialectes alémaniques ou n'ont pas su adapter suffisamment les moyens linguistiques employés en fonction des possibilités de leurs vis-à-vis.

Les interlocuteurs ont au préalable été formés pendant deux jours. Ils ont exécuté des versions précédentes des tests avec des élèves de 6^e et de 9^e année, et ont été filmés pendant l'exercice. Ils ont reçu une introduction à leur travail au moyen d'analyses des enregistrements vidéo et de commentaires ouverts. Parmi les priorités de la formation, citons, outre une introduction générale au dispositif de test, une sensibilisation des interlocuteurs à la nécessité de mener des entretiens comparables et de donner à la conversation une orientation claire, mais aussi naturelle que possible.

3.7.3 Collecte des données

Les tests oraux se sont déroulés dans les écoles des participants, ce qui a nécessité une souplesse considérable de la part de l'équipe d'examineurs, composée à chaque fois de trois personnes – des démarches qui ne seraient guère envisageables avec des échantillons représentatifs sur le plan statistique. Tous les tests ont été filmés.

Chaque élève devait faire toutes les parties des tests:

- prise de contact à trois;
- production orale;
- interaction orale avec un partenaire de conversation du même âge;
- interaction orale avec un interlocuteur adulte.

Du fait qu'il y avait un nombre assez élevé d'exercices différents, chaque exercice n'a été fait que relativement peu de fois dans chacune des trois langues. Les exercices d'allemand ou de français qui ont été réalisés uniquement par des élèves de sixième ou par des élèves de neuvième année ont été exécutés par 8 élèves; les exercices proposés dans plus d'un degré et les exercices d'anglais, par 14 élèves. En tout, nous avons obtenu un total de 540 performances orales.

Des indications temporelles étaient seulement données pour la préparation. Avant les parties «production orale» et «interaction orale» avec un élève du même âge, les élèves disposaient, après une première lecture de l'exercice écrit, d'une minute pour se préparer mentalement au thème proposé, activer le bon champ lexical et noter quelques mots clés (rien de plus!). Ce temps de préparation devait donner la possibilité aux élèves de se placer dans le contexte opérationnel proposé et d'activer leurs ressources en matière de communication. Ces indications ont été données dans le but de permettre un rapprochement entre les conditions du test et des situations de conversation semblables survenant en dehors de toute situation de test. Avant les deux autres parties, aucune préparation n'était nécessaire, puisque la planification de la conversation était en principe assumée par l'interlocuteur et que le but était de communiquer aussi spontanément que possible. En ce qui concerne le temps de parole pour les différents exercices, il n'y avait aucune indication précise. On a, en règle générale, laissé les élèves parler aussi longtemps qu'ils s'exprimaient en rapport avec l'exercice. Les transitions entre les différentes parties du test ont été réalisées de façon aussi naturelle que possible. Le temps de test s'est élevé, en tout, à près de 20 minutes par élève.

3.7.4 Evaluation

Chaque test a été évalué en trois étapes distinctes: 1) exercice relatif à la production orale; 2) exercice relatif à l'interaction entre pairs (*peer interaction* – l'accent étant mis, à chaque processus d'évaluation, sur un seul élève); 3) entretien (y compr. prise de contact tout au début).

Les parties de test 1) et 2) relatives à la production et à l'interaction orales ont été évaluées de manière très détaillée. Pour ce faire, nous avons employé des instruments qui ont été développés spécialement à cet effet au sein du consortium. A l'aide des catégories retenues, on cherche à mettre en œuvre, de manière assez complète, les catégories pertinentes pour l'approche orientée sur l'action du *Cadre de référence*. Bon nombre de ces descriptions reprennent des catégories et des éléments de description du *Cadre de référence*. On ne vise toutefois pas une mise en regard directe avec le *Cadre de référence*; il s'agit plutôt d'obtenir une description de la performance qui fasse abstraction, par exemple, des exigences liées aux différents exercices – comme cela a d'ailleurs été fait pour les exercices d'écriture. Les critères auxquels nous avons eu recours sont tantôt spécifiques à l'exercice (conceptuels-thématiques), tantôt propres au type de discours en

question (par ex.: qualité de l'acte d'information ou d'argumentation); parfois, ils sont liés à l'aptitude considérée (capacité d'interaction, capacité de production) ou encore se rapportent de manière générale à toutes les performances orales réalisées (éléments linguistiques et certains aspects pragmatiques). Ils ne se réfèrent toutefois jamais uniquement à l'une des trois langues testées. Les échelles pour les critères analytiques comportent à chaque fois des barèmes allant de 0 à 4 ou 5, ce qui permet de saisir un large spectre de compétences et d'exclure ainsi dans une grande mesure des «effets plancher» (*floor effects*) ou des «effets plafond» (*ceiling effects*), c'est-à-dire que l'on peut encore opérer des différenciations dans les performances qui se classent aux extrémités de l'échelle. Plusieurs exercices proposaient deux parties avec des types de discours différents, par ex. «informer / décrire» en premier, puis «exprimer son opinion / argumenter». Dans ces cas-là, les deux parties ont été évaluées séparément, parce que l'on peut en principe s'attendre à ce que les différentes exigences liées à un type de discours donné auront des conséquences sur la façon dont les ressources linguistiques pourront être mobilisées.

L'imbrication des critères au-delà des limites des exercices et des aptitudes sera brièvement démontrée ici à l'aide de deux exercices d'interaction orale et de deux exercices de production orale:

<i>IP2 «Camp scolaire»</i>	<i>IP5 «L'uniforme à l'école?» (Partie II)</i>	<i>P2 «Portrait de classe»</i>	<i>P5 «TV – ordinateur – téléphone?»</i>
Capacité de production / compétence relative au discours (2 éléments)		Capacité d'interaction (8 éléments)	
Evaluation globale de l'exécution de l'exercice (type «informer / décrire»)	Evaluation globale de l'exécution de l'exercice (type «exprimer son opinion / argumenter»)	Evaluation globale de l'exécution de l'exercice (type «informer / décrire»)	Evaluation globale de l'exécution de l'exercice (type «informer / décrire»)
Réalisation des lignes directrices de l'exercice IP2	Réalisation des lignes directrices de l'exercice IP5	Réalisation des lignes directrices de l'exercice P2	Réalisation des lignes directrices de l'exercice P5
Aspects linguistiques et pragmatiques: Prononciation / intonation Aisance Cohérence et cohésion Grammaire (domaine verbal: 3 éléments; domaine de la phrase: 3 éléments) Vocabulaire (2 éléments)			

Ainsi pour toutes les performances relevant de la production orale on applique les mêmes critères de «capacité de production»; il en va de même pour l'interaction orale. Par contre, l'évaluation globale de l'exécution des exercices est différente d'un exercice à l'autre, sauf lorsque ces exercices nécessitent le même type de discours. Quant à la réalisation des lignes directrices, elle est jugée de manière spécifique à chaque exercice.

Pour la troisième partie du test (entretien et prise de contact), l'intention et le procédé de l'évaluation ont été fondamentalement différents: les performances des élèves devaient être classées directement en fonction des niveaux de référence, au moyen de critères tirés du *Cadre de référence* de même que d'autres critères attribués au *Cadre de référence*. Un classement qui devait du reste s'avérer possible et approprié, avec ce dispositif de test, puisque les interlocuteurs avaient pour tâche de sonder, au fil de la conversation, les limites de la capacité de performance langagière des élèves testés.

Dans le cadre de l'évaluation qui a été effectuée, on a procédé, dans un premier temps, à une évaluation globale, de type holistique, de l'acte de parole. L'instrument utilisé à cet effet correspond largement à l'échelle globale pour la communication orale (cf. Schneider / North 2000, 341) du projet de recherche suisse, d'où sont tirés la plupart des descripteurs du *Cadre de*

référence, avec en plus quelques compléments du projet *IEF / lingualevel* (pour le domaine de niveaux inférieur). Cette estimation permet d'établir une première conclusion sur l'aisance d'un élève dans ses actes de parole. Dans le cadre de l'évaluation analytique qui a suivi, différents aspects qualitatifs de l'acte langagier ont été étudiés séparément. L'instrument employé est étroitement calqué sur le tableau 3 du *Cadre de référence* (Conseil de l'Europe 2000, 28 sq.) et, comme l'échelle globale, complété par *IEF / lingualevel*; cette échelle comporte les aspects qualitatifs suivants: (capacité d')interaction, étendue (du vocabulaire), correction (grammaticale), aisance, prononciation / intonation et cohérence (premièrement dans le sens de la cohésion).

Les évaluations ont été effectuées, pour chaque langue, par deux étudiantes / étudiants en linguistique avancés (une doctorante, cinq licenciées / licenciés). Parmi eux, deux étaient de langue maternelle allemande, une de langue maternelle français, et les trois autres disposent d'une compétence linguistique élevée et d'une bonne expérience de la langue.

Nous avons procédé en six étapes pour permettre aux évaluateurs de se familiariser avec les instruments et de s'aligner fondamentalement sur les niveaux (ou plus précisément: les niveaux affinis A1.1, A1.2, etc.) du *Cadre de référence*, tant pour l'interaction que pour la production:

1. **Informations** sur HarmoS, notamment sur la compréhension des compétences; référence au *Cadre européen commun de référence*;
2. **Explication des exercices**, l'accent étant mis sur le déroulement, les contenus et les objectifs;
3. **Auto-évaluation de sa propre compétence orale** dans une langue étrangère de son choix, à l'aide de l'échelle globale relative à l'expression orale ainsi que des descripteurs tirés du *Portfolio européen des langues pour jeunes et adultes* (Schneider / North / Koch 2000).
4. **Familiarisation concrète** avec l'échelle globale pour la codification de la partie 3 du test, interaction avec des locuteurs natifs: les évaluateurs classent les différentes descriptions par niveau et font ensuite ressortir, dans la grille d'évaluation analytique détaillée relative à l'interaction, les caractéristiques des différentes descriptions de niveaux; ils soulignent les points importants et approfondissent leur compréhension des niveaux en en discutant avec les autres évaluateurs et les formateurs.
5. **Evaluation commune** de six entretiens (deux par langue) dans le groupe au complet. On procède à une évaluation globale après la première moitié environ de chaque conversation; on passe alors à l'évaluation détaillée grâce aux critères analytiques.
6. **Evaluation individuelle** de trois autres exemples (un par langue), avec comparaison finale et discussion des résultats.

Afin d'assurer la comparabilité des évaluations dans les différentes langues, trois performances d'élèves (une par langue) étaient évaluées conjointement au début de chaque journée d'évaluation. Chacun procédait d'abord à une évaluation individuelle, puis tous en parlaient jusqu'à ce qu'un consensus soit trouvé. Une autre mesure prise pour garantir la fiabilité inter-examineur a été de faire en plus évaluer par une experte du consortium un échantillon de 36 performances orales. Lorsqu'il y avait divergence, les personnes concernées parlaient ensemble de l'évaluation. La fiabilité inter-examineur a aussi pu être vérifiée en permanence grâce aux 10% environ de performances qui étaient à chaque fois évaluées par les deux évaluatrices / évaluateurs compétents pour une même langue cible. En vue d'assurer une meilleure fiabilité intra-examineur, la dernière performance évaluée le jour précédent était à chaque fois reprise, au début de la matinée, par le même évaluateur, qui l'examinait à nouveau; enfin, les évaluations ont été comparées et, quand cela s'est avéré nécessaire, on a procédé aux corrections qui s'imposaient. Après la pause de midi, les deux évaluatrices / évaluateurs travaillaient tous deux sur la même performance, comparaient leurs évaluations et passaient en revue les différences.

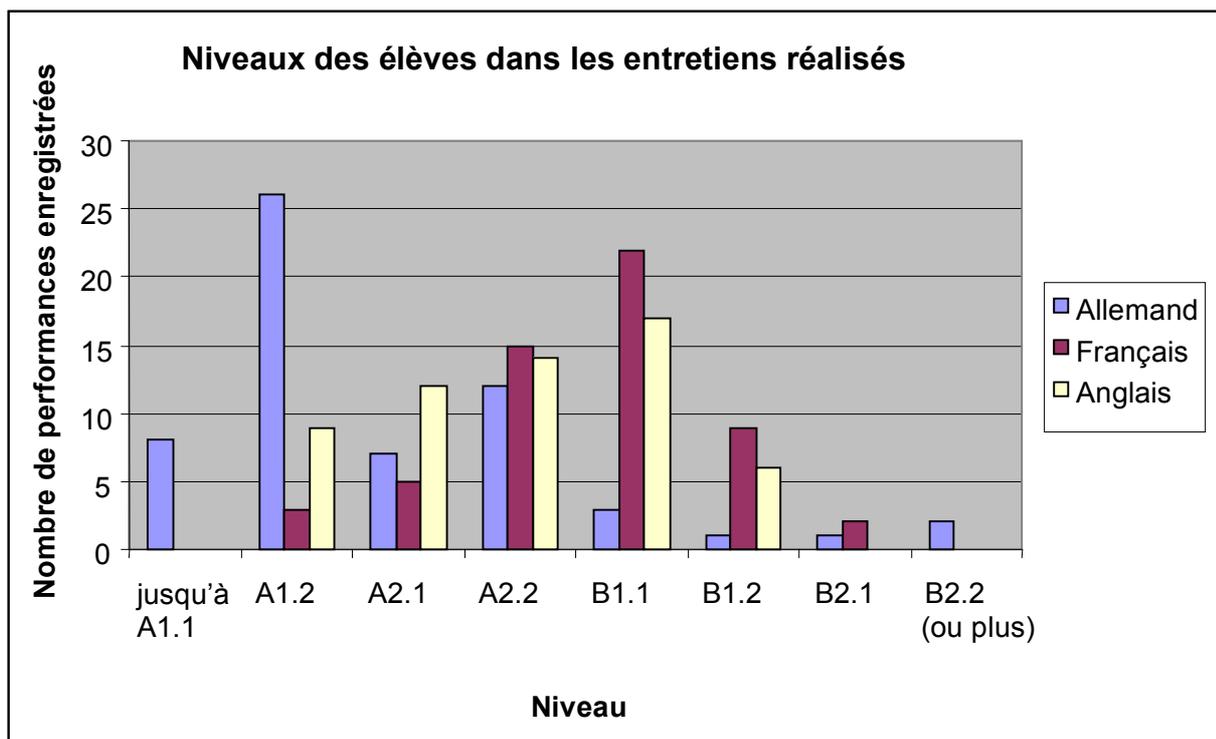
3.7.5 *Analyses et résultats*

Seules les données d'évaluation relatives à la partie 3 du test ont pu être valablement exploitées sur le plan statistique; il s'agit de l'entretien réalisé avec un partenaire de conversation adulte, dont la langue maternelle est la langue testée. Cependant, l'élaboration des exercices et des instruments d'évaluation, la formation des interlocuteurs, l'exécution des tests et l'évaluation elle-même ont livré un grand nombre de conclusions importantes, sur lesquelles il sera possible de

construire à l'avenir. Qui plus est, le projet a permis la réalisation d'environ 50 heures de matériel filmé ainsi qu'une série de transcriptions scientifiques de performances d'élèves, qui peuvent être mises à profit pour des analyses subséquentes.

Les données relatives à la partie 3 ont été exploitées en premier lieu avec le logiciel de statistiques *Facets*: on a opté pour une analyse *Rasch* à plusieurs facettes, qui permet de compenser particulièrement bien la différence de sévérité entre les évaluateurs et de vérifier les propriétés psychométriques des critères. Les résultats obtenus sur les critères confirment en partie ce qui a déjà été révélé par des enquêtes antérieures; mais ils posent aussi de nouvelles questions: comme nous l'avons constaté au fil de l'évaluation déjà, le critère de la prononciation / intonation contribue peu à une différenciation des élèves. Ces derniers sont presque tous assez bons en ce qui concerne ce paramètre; de plus, ceux qui rencontrent d'habitude des difficultés ne sont pas forcément moins bons que les autres sur ce point. Il n'est donc pas surprenant que cette variable ait dû être exclue sur la base des indices psychométriques. Comme toujours dans pareils cas, il faut décider si – et si oui, sous quelle forme – cet aspect important, surtout aux yeux des non-professionnels, doit être inclus dans de futures évaluations. Vu qu'il est visiblement presque évident que, chez les élèves actuels, l'accent n'empêche guère le locuteur de se faire comprendre, ce critère pourrait bien être entièrement écarté. Par contre, il a été plutôt surprenant de constater, au cours de l'analyse des critères d'évaluation, que celui de la «correction (grammaticale)» n'a lui non plus pas obtenu de bonnes valeurs, du moins en allemand et en anglais. Apparemment, le degré d'élaboration ou de correction grammaticale n'est pas fondamentalement important pour réaliser une communication efficace; il se peut aussi que ces résultats révèlent que de se concentrer sur des aspects grammaticaux risque plutôt de gêner la performance de communication générale (par ex. dans les domaines de l'aisance et de la capacité d'interaction). La question des conséquences sur les futurs instruments d'évaluation se pose donc aussi avec ce critère. Si le domaine de la grammaire est séparé des autres points et forme une dimension en soi, il devrait pouvoir être mieux saisi, de façon plus variée. D'ailleurs, les difficultés, exprimées de manière informelle, qu'ont rencontrées les évaluatrices / évaluateurs dans ce domaine indiquent qu'il est de toute façon nécessaire de faire quelque chose par rapport à cet état de fait. Mieux saisir ces critères, de façon plus précise – et peut-être de manière spécifique à chaque langue – pour les différents niveaux est un travail qui devrait être entrepris dans un avenir proche, par exemple dans le cadre de travaux de séminaires ou de master.

Parmi les objectifs premiers de l'enquête relative à l'expression orale figurait tout simplement le projet de créer une série de performances orales d'élèves, qui puissent être utilisées comme illustrations et base de discussion pour le débat portant sur les standards (et sur les plans d'études). Le graphique suivant donne un aperçu du matériel disponible dans chaque langue:



III. 22: Aperçu des niveaux des élèves ayant passé des tests d'expression orale dans le cadre de l'enquête HarmoS (sur la base des évaluations brutes, et non pas de l'échelle de Rasch).

Il apparaît que le spectre de niveaux souhaité n'a pas pu être atteint pour chaque langue. Pour l'allemand en langue étrangère, on a filmé un grand nombre de performances très modestes dans les quelques classes choisies de façon assez aléatoire, surtout dans les classes de sixième année. En revanche, pour le français, on a obtenu des performances excellentes, mais pas représentatives; elles ont été réalisées par des élèves de sixième année de la partie germanophone du canton de Fribourg, si bien que là on constate plutôt des lacunes dans le domaine de niveaux inférieur. Mais dans l'ensemble, c'est un beau corpus qui a été constitué, et qui peut être exploité de bien des manières. Quant à savoir combien les niveaux proposés comme standards de base sont réalistes, il n'est pas possible de trancher la question de façon définitive sur la base des résultats obtenus en expression orale. Néanmoins, le fait est que, parmi les élèves testés, il s'en trouve un nombre élevé qui n'atteindraient pas les standards de base de leur niveau⁷⁴.

Comme cela a déjà été mentionné, la troisième partie du test a été réalisée dans le but de faire apparaître, grâce au comportement adaptatif des interlocuteurs, le «vrai» niveau de compétence des élèves. Et, de l'avis des personnes impliquées, cela s'est fait dans bien des cas, même si les interlocuteurs ne se sont évidemment pas toujours comportés de façon idéale – et ce, justement parce qu'ils disposaient d'une grande marge de manœuvre et qu'ils devaient faire preuve de souplesse. Il est par ailleurs certain que le facteur «interlocuteur» entraîne aussi, jusqu'à un certain point, un effet de distorsion dans ce contexte, parce que des comportements sous-jacents comme la bonne volonté ou la critique ne peuvent jamais être complètement contrôlés. A cela s'ajoute le fait que certains élèves – par exemple les élèves de sixième année de Fribourg – n'ont guère eu de peine à communiquer authentiquement avec des locuteurs de la langue cible, parce qu'ils bénéficiaient d'une grande expérience dans ce domaine, alors que d'autres ne pouvaient visiblement pas s'appuyer sur des expériences du quotidien (même extrascolaires) et que, de ce fait, c'était pour eux un défi bien plus grand sur le plan émotionnel.

Plus que dans la troisième partie, c'est en considérant les performances réalisées dans la deuxième partie du test (*peer interaction*) que l'on a l'impression que les élèves n'ont que rarement pu ou voulu montrer ce dont ils seraient probablement capables s'ils en avaient l'occasion ou s'ils

⁷⁴ Classe de 8^e année (actuelle 6^e année), 1^{re} langue étrangère: A2.1; classe de 8^e année, 2^e langue étrangère: A1.2; classe de 11^e année, 1^{re} ou 2^e langue étrangère: A2.2.

étaient explicitement invités à agir. A l'évidence, un grand nombre d'élèves n'avait pas l'habitude de faire des tests d'interaction avec un partenaire, ou même de communiquer un contenu (plutôt que des jeux de rôle ou des exercices de structure) à un élève du même âge, dans la langue étrangère. On a notamment observé qu'un nombre relativement élevé d'élèves avait tendance à se cacher plutôt qu'à se montrer, si bien qu'au final, nous ne disposions (souvent pour l'un des deux élèves) que de très peu de matériel pour tirer des conclusions sur la capacité de communication de l'élève en question. Les exercices dans lesquels il fallait manier du matériel (surtout l'exercice IP1) ont aussi joué un rôle certain dans le fait que les élèves se sont parfois peu exprimés dans la partie *peer interaction*. Une bonne partie pouvait assez bien être réalisée par un langage non verbal ou avec peu de mots, que les élèves répétaient souvent. – L'interaction entre pairs, dite *peer interaction*, s'est développée ces dernières années jusqu'à devenir un élément standard de la partie «expression orale» des examens de diplômes internationaux connus, et elle est généralement reconnue comme y ayant sa place, ce qui a pour effet en retour (*washback effect*) positif le fait que l'expression interactive est effectivement entraînée, par deux, dans le cadre des préparations aux examens. Et justement, dans des classes comptant 20 élèves ou plus, cette forme sociale est l'une des rares possibilités de pratiquer, mais aussi d'évaluer, régulièrement la capacité d'interaction des élèves. Par conséquent, les expériences d'HarmoS, qui semblent probablement parfois un peu teintées de scepticisme, ne devraient pas amener le lecteur à conclure que cette forme de test n'est pas appropriée, mais plutôt que cette forme de collaboration devrait réellement être encouragée dans l'enseignement oral des langues étrangères.

Les données relatives à la production orale n'ont, comme celles qui portent sur l'interaction entre pairs (*peer interaction*), pas pu être mises en échelle à cause d'un volume de données insuffisant. Toutefois, dans l'optique d'illustrer les propositions de standards de base, plus de 50 performances d'élèves en tout, relatives à la production orale, ont été classées conjointement par des expertes et des experts du consortium et attribuées aux niveaux de référence⁷⁵. Une comparaison simple de ce classement avec les attributions de niveaux effectuées sur la base de la partie 3 du test révèle une tendance à l'«effet seuil»: les élèves plus faibles, dans le domaine de niveaux A1.2 / A2.1, ont pu réaliser une meilleure performance (ou faire croire à une plus grande compétence de communication langagière) dans l'interaction avec l'interlocuteur adulte que dans la production orale, où ils «crochaient» souvent et avaient beaucoup de peine à s'exprimer sans le soutien actif d'une personne plus compétente qu'eux au niveau du langage.

Comme cela a été mentionné, les instruments d'évaluation complets qui ont servi à codifier les performances dans les parties de tests 1 et 2 n'ont pas pu, par manque de données, être vérifiés et étalonnés en fonction de leurs propriétés statistiques. Cependant, dans le cadre de la collaboration intensive et de très bonne qualité qui s'est établie avec les évaluateurs, il a quand même été possible d'optimiser de nombreuses formulations et de récolter des commentaires de toute sorte sur les critères employés. Des éléments de ces formulations ont été intégrés dans la description des niveaux de compétences, lorsqu'ils caractérisaient des phénomènes intervenant de manière répétée dans les performances des différents niveaux.

3.7.6 Perspectives: examen de la compétence orale dans le cadre du monitoring de l'éducation

Dans le cadre de la scolarité obligatoire, l'expression orale relève sans nul doute des compétences de communication langagières centrales. En même temps, les tests classiques d'expression orale nécessitent – à l'instar de ceux qui ont été décrits ci-dessus – beaucoup plus d'efforts pour leur exécution que les tests portant sur les aptitudes langagières de réception, en particulier. Et pourtant, il serait problématique, du fait du signal que cela communiquerait, que l'expression orale ne trouve pas, pour des raisons de coûts, sa place dans un futur monitoring de l'éducation.

L'institut IQB de Berlin développe, sur mandat de la conférence allemande KMK, des tests d'expression orale sur ordinateur. Le but visé par des tests sur ordinateur de ce genre est de créer

⁷⁵ Dans l'ensemble, il a été difficile d'attribuer, avec une certaine assurance, l'un des niveaux de référence aux élèves, particulièrement dans les niveaux inférieurs, sur la base de leur réussite (ou de leur échec) dans un seul exercice de production orale. Un «style de communication» particulièrement dirigé par la volonté d'éviter toute erreur chez un certain nombre d'élèves plus faibles (surtout des élèves romands de sixième année) a rendu la tâche des évaluateurs extrêmement difficile; en fait, il n'a guère été possible d'établir une évaluation équitable.

un procédé dans lequel l'exécution et l'évaluation s'effectuent complètement automatiquement; on obtient ainsi des tests qui sont très bon marché. Mais il y a là aussi un prix à payer puisqu'en règle générale cela pose des problèmes de validité, dans le sens où les conversations sont réduites, jusqu'à un certain point, à des successions de *prompts* de l'ordinateur et de réaction des élèves. Les exigences se situent plutôt au niveau des moyens linguistiques que de la capacité d'interaction. Le potentiel des tests sur ordinateur dans l'expression orale dépend fortement de l'état de la technologie dans le domaine du traitement des langues naturelles (*natural language processing*); cela se vérifie tant dans l'authenticité et le naturel de l'interaction que dans la fiabilité de l'évaluation. Les progrès réalisés dans ce domaine sont considérables; les rapports sur les tests d'expression orale (mais aussi d'expression écrite) sur ordinateur ne sont plus seulement négatifs, si bien qu'on observe aujourd'hui un potentiel qui devrait dans tous les cas être examiné de manière approfondie, d'autant plus qu'envisager un monitoring de l'éducation sans tests d'expression orale n'est pas la bonne solution dans l'alternative qui se pose à l'heure actuelle.

Une réflexion s'impose, au-delà du monitoring de l'éducation proprement dit, face à la question suivante: dans quelle mesure d'autres moyens d'assurance qualité pourraient décharger quelque peu le monitoring de l'éducation et contribuer à promouvoir la compétence langagière orale? Une approche plutôt informelle pourrait être proposée avec la mise à disposition d'instruments d'évaluation tels que *lingualevel*, accompagnée d'une introduction active à ces instruments; elle serait proposée dans le cadre de la classe et permettrait de mesurer les performances des élèves à l'aide d'échelles externes qui facilitent la comparaison. Une autre possibilité serait de proposer des certificats de langues dans les écoles publiques. Des parties orales bien développées font partie intégrante, depuis longtemps, des examens de certificats, et peuvent visiblement aussi être financées⁷⁶. Des tests informels assez ciblés, renforcés par une offre d'examens de certificats comprenant des parties orales complètes, permettraient jusqu'à un certain point de renoncer à des solutions maximales dans le cadre du monitoring de l'éducation.

4 Domaines de compétences n'ayant pas été étudiés empiriquement

4.1 Médiation linguistique

Le *Cadre de référence* distingue, en plus de la réception, de la production et de l'interaction, orales et écrites, un quatrième domaine de compétences: la médiation linguistique orale et écrite. Les activités orales et écrites relevant de la médiation linguistique permettent à des personnes qui ne communiquent pas directement dans la même langue de se comprendre. Font partie des activités langagières de médiation la traduction et l'interprétation (dans notre contexte, il ne s'agit toutefois pas de traduction ou d'interprétation au sens littéral), ainsi que le fait de résumer ou de paraphraser des textes dans la même langue ou dans une autre langue. Les activités langagières de médiation consistent à retransmettre, oralement ou par écrit, pour une personne qui ne comprend pas ou pas suffisamment la langue utilisée, le sens de communications que l'on a entendues ou lues, dans une autre langue qu'elle comprendra.

Le *Cadre de référence* précise que: «les activités langagières de médiation, (re)traitant un texte déjà là, tiennent une place considérable dans le fonctionnement langagier ordinaire de nos sociétés» (CECR, 18). Par conséquent, on part du principe que les activités langagières de médiation devraient aussi faire partie des domaines de l'évaluation et de la certification. L'évaluation ne se concentrerait pas sur la compétence du sujet dans *une seule* langue, mais porterait sur les «capacités de gestion d'un tel répertoire» comprenant différentes langues (CECR, 130).

S'alignant sur le *Cadre de référence*, le consortium Langues étrangères a adopté la médiation linguistique comme l'un des domaines de compétences s'intégrant dans le modèle de compétences. Ce choix sort en fait du cadre du mandat reçu puisque l'élaboration de standards de base est prévue, dans le projet HarmoS, pour des disciplines isolées. En effet, le mandat consistait

⁷⁶ Dans le cadre de l'école publique, il peut toutefois s'avérer très problématique que des mesures essentielles d'assurance qualité deviennent du ressort de la volonté personnelle, donc de l'économie privée, et doivent être financées par les parents des élèves.

– malheureusement, mais de façon réaliste vu les conditions temporelles dans lesquelles s’est déroulé le projet – à concevoir un modèle de compétences et des standards pour les disciplines de l’allemand, du français et de l’anglais, donc pour des disciplines isolées, et non pas un modèle qui s’étendrait à une compétence plurilingue.

A l’heure actuelle, la médiation linguistique ne figure que dans quelques rares plans d’études et c’est à peine si les activités langagières de médiation jouent un rôle actuellement dans l’enseignement des langues étrangères. Pourtant, on doit s’attendre à ce que la médiation linguistique prenne plus de place si l’apprentissage simultané de deux langues étrangères ainsi que – nous l’espérons – une didactique intégrée des langues correspondante sont mis en œuvre dans le cadre du scénario 2010-2012.

En vue de cette évolution, nous souhaitons esquisser des pistes, du moins dans les grandes lignes, afin de montrer dans quelle direction on pourrait se diriger pour la formulation de standards dans ce domaine.

Dans le cadre du présent projet, il n’a pas été possible de mener des enquêtes empiriques sur la médiation linguistique. Du reste, si l’on avait réalisé, à ce stade, des tests sur les compétences en médiation linguistique, les résultats n’auraient pas été très probants, vu que la médiation linguistique n’est guère pratiquée et entraînée actuellement dans le cadre de l’enseignement des langues étrangères.

Les descriptions de niveaux qui ont été intégrées dans le document relatif aux standards de base ne reposent donc pas – comme d’ailleurs toutes les descriptions proposées jusqu’ici en rapport avec la médiation linguistique – sur des résultats empiriques, qui auraient été obtenus à partir d’exercices de test spécifiques à la médiation linguistique, mais sont dérivées d’autres sources, qui seront brièvement exposées ci-dessous.

Le *Cadre de référence* ne contient pas de descriptions de compétences pour les activités langagières de médiation. Certes, on trouve quelques rares descriptions de «capacités à faire» («Kannbeschreibungen») pour la médiation linguistique parmi les standards de formation publiés par la Conférence allemande des ministres de l’éducation et des affaires culturelles des länder (KMK), de même que dans les standards de formation et les programmes fondamentaux de différents länder allemands. Cependant, ces descriptions sont souvent très générales. Une liste systématique, assez détaillée, de descriptions de «capacités à faire» relevant de la médiation linguistique figure dans *Profile deutsch* (Glaboniat et al. 2005). En plus des descriptions de «capacités à faire», on y trouve systématiquement des exemples d’activités langagières de médiation dans divers domaines de la vie (adulte), notamment dans la sphère privée et publique, ainsi que dans les domaines de la formation et de la profession. Mais *Profile deutsch* ne propose pas de mise en œuvre pour les descriptions de «capacités à faire», et ne comporte pas d’exemples d’exercices concrets. En outre, les descriptions de «capacités à faire» n’ont pas fait, en tant que telles, l’objet d’une validation empirique pour l’instant.

Les descriptions de «capacités à faire» de ces sources pourraient être utilisées comme impulsion de départ, mais elles ne pourraient pas être reprises; celles des standards de formation de la KMK ne s’y prêtent pas parce qu’elles sont trop vagues et pas assez spécifiques à des niveaux, celles de la liste complète de *Profile deutsch* pas non plus parce qu’elles sont plutôt destinées à des adultes et que la systématique qui y a été adoptée entraîne des formulations assez compliquées.

Voici comment nous avons procédé pour la sélection et la formulation des activités langagières de médiation.

Dans une première étape, nous avons restreint le champ des exercices, en excluant la médiation linguistique qui se produit au sein d’une seule et même langue, comme le fait de résumer un texte compliqué en le simplifiant, par ex. quand on paraphrase des textes spécialisés pour des profanes. Nous nous limitons donc à la médiation linguistique entre une langue étrangère et la langue scolaire locale premièrement et, secondement, entre une première et une seconde langues étrangères.

Dans une deuxième étape, nous avons identifié, dans le champ ainsi délimité, des types d’activités langagières de médiation. Ces activités résultent de la combinaison des facteurs: 1. langues

impliquées; 2. oral-écrit; 3. type de situation de communication: directe (en face à face) ou décalée dans le temps.

Nous avons sélectionné, sans prétention d'exhaustivité, les combinaisons suivantes, qui nous paraissent pertinentes pour le groupe cible:

- a) Médiation linguistique oral ↔ oral: alternativement de la langue étrangère dans la langue scolaire locale et de la langue scolaire locale dans la langue étrangère (interprétation informelle);
- b) Médiation linguistique oral ↔ oral: alternativement d'une langue étrangère dans une deuxième langue étrangère (interprétation informelle);
- c) Médiation linguistique oral ↔ oral: de la langue scolaire locale dans la langue étrangère;
- d) Médiation linguistique oral → écrit: de la langue scolaire locale dans la langue étrangère;
- e) Médiation linguistique oral → oral / écrit: de la langue étrangère dans la langue scolaire locale;
- f) Médiation linguistique écrit → oral / écrit: de la langue étrangère dans la langue scolaire locale;
- g) Médiation linguistique écrit → écrit: d'une langue étrangère dans une autre.**

Dans une troisième étape, nous avons formulé des descriptions de «capacités à faire» qui s'orientent, pour ce qui est du potentiel d'action dont dispose une personne dans des activités de médiation linguistique, selon les compétences qu'elle a acquises dans les aptitudes langagières qui interviennent, à savoir celles de réception, de production et d'interaction. Les exigences seront différentes suivant si la retransmission s'effectue oralement ou sous forme écrite, étant donné que les standards de base pour le domaine de l'expression orale ont été fixés plus haut que pour l'écrit. Dans les cas où quelque chose a été lu ou entendu dans une langue étrangère et est retransmis dans la langue scolaire locale, on applique les standards de base pour la compréhension de l'oral ou de l'écrit dans la langue étrangère respective.

C'est la raison pour laquelle les formulations des «capacités à faire» relatives à la médiation linguistique reprennent littéralement des éléments de descriptions des autres domaines, ce qui permet de donner des précisions et des restrictions propres aux différents niveaux (par ex. «thèmes bien connus», «à l'aide de phrases et de tournures simples», etc.).

Les descriptions de «capacités à faire» (en gras) sont à chaque fois quelque peu concrétisées au travers de divers exemples d'activités langagières de médiation dans des situations du quotidien et de la vie scolaire.

Pour ce qui est de la qualité des productions orales ou écrites dans la langue étrangère, on pourrait, pour compléter, se référer aux descriptions relatives au spectre de moyens linguistiques à disposition, à l'aisance et à la correction, qui sont notées pour l'expression orale et écrite du standard de base en question. A noter toutefois que le recours aux descriptions de compétences des diverses aptitudes orales et écrites de réception et de production revêtent un caractère hypothétique. Il s'agirait de vérifier empiriquement, dans le cadre d'enquêtes complémentaires, si – et si oui, dans quelle mesure – par ex. la capacité de traitement limitée lorsqu'il faut agir dans deux langues en même temps ou la nécessité d'être attentif aux besoins des partenaires de communication permettent encore des performances de même nature, dans les diverses combinaisons relevant de la médiation linguistique, que celles décrites pour les différentes aptitudes impliquées.

Les descriptions de «capacités à faire» formulées, à titre d'exemple, pour des classes de 11^e année et les exemples d'activités langagières de médiation peuvent d'une part donner des impulsions pour une plus grande prise en compte de ce domaine de compétences dans les plans d'études et les ouvrages d'enseignement. D'autre part, elles peuvent aussi servir de point de départ pour l'élaboration d'exercices dans ce domaine. Car si un jour le scénario 2010-2012 pour la réforme de l'enseignement des langues étrangères est réalisé et qu'une didactique intégrée des langues est mise en œuvre, il faudra développer, dans le cadre d'une enquête subséquente dans la lignée du projet HarmoS, des exercices de tests portant sur les combinaisons de la médiation linguistique, et formuler, pour ce domaine de compétences, des standards basés sur les exercices, en prenant appui sur une base empirique.

4.2 Compétences transversales

4.2.1 Généralités

La compétence interculturelle et la compétence méthodologique figurent parmi les composantes des «compétences générales» dans le modèle de compétences «Langues étrangères» – tout comme dans le modèle dont il s'inspire, celui du *Cadre de référence*. Si l'importance de ces domaines de compétences est expressément soulignée, il manque néanmoins des descriptions de compétences et des échelles pour ces domaines-là. Le *Cadre de référence* a régulièrement fait l'objet de critiques sur ce point, et les auteurs tant du *Cadre de référence* que de projets réalisés par la suite dans la périphérie du *Cadre de référence* sont conscients de ces lacunes. C'est ainsi par exemple que Daniel Coste a constaté que «tout ce qui touche à la compétence (inter)culturelle reste peu développé dans le *Cadre* et demande là aussi, dans bien des contextes, des descripteurs, voire des types de descripteurs, autres que ceux dont on dispose déjà» (Coste 2007, 51). Il est compréhensible que certains praticiens, et surtout les auteurs de plans d'études et les enseignants, attendent aussi du projet «HarmoS - Langues étrangères» des propositions de standards pour ces domaines de formation importants. Cependant, il n'est pas possible, pour plusieurs raisons, de répondre à cette attente dans le cadre du projet «HarmoS - Langues étrangères».

Dès le départ, le consortium Langues étrangères a souligné qu'il reste encore beaucoup de travaux fondamentaux à réaliser pour pouvoir parvenir à des définitions et à des niveaux de compétences étayés sur le plan théorique et empirique. Il a également relevé que des projets de recherche supplémentaires seraient nécessaires à cet effet (par ex. dans le cadre du programme national de recherche, PNR 56, «Diversité des langues et compétences linguistiques en Suisse»). De tels projets n'ont cependant pas été lancés.

Dans la lignée de l'expertise Klieme, HarmoS se concentre sur la question des compétences dans des disciplines isolées, en l'occurrence les compétences des élèves dans certaines langues étrangères, et non pas sur des compétences interdisciplinaires. Les compétences interculturelles et méthodologiques, de même que la conscience linguistique sont très proches des disciplines que sont les langues étrangères, mais elles ne peuvent se développer dans le cadre d'une seule discipline de langues étrangères, ni d'ailleurs dans les langues étrangères uniquement.

Vu le calendrier serré du projet HarmoS, il n'a pas non plus été possible de tester des exercices ou des questionnaires sur des questions d'attitudes par exemple, de comportement adopté en vue de résoudre des problèmes, ou encore de savoir culturel. De notre avis, il serait problématique de formuler, sans aucune base empirique, des descriptions étalonnées de compétences et de vouloir proposer des standards de base. On courrait alors le risque que la différence de statut entre les descriptions des compétences communicatives langagières d'un côté et, de l'autre, celles des compétences interculturelles et méthodologiques ne soit pas suffisamment prise en considération dans la pratique.

Nous nous voyons confortés dans notre analyse lorsque nous lisons par exemple dans le manuel présentant les standards de formation autrichiens, qui comporte une liste de capacités pour ces domaines de compétences: «les descripteurs relatifs aux domaines de capacité dynamiques que sont la *compétence communicationnelle, interculturelle, sociale* et les *stratégies d'apprentissage* ne peuvent que très difficilement être classés en fonction des différents niveaux; ils se comprennent donc, du moins provisoirement, comme s'appliquant à plusieurs niveaux⁷⁷». Et puis, plus loin: «Les descripteurs des *capacités dynamiques* risquent – et les auteurs en ont tout à fait conscience – de dériver vers des exigences exagérées de «fitness» interculturel, social et en matière de stratégies d'apprentissage. Pourtant, ils ont leur justification, pas comme des objectifs d'enseignement à atteindre à tout prix, mais comme des lignes de pensée et d'action, qui ne peuvent certes pas vraiment être catégorisées et mesurées en fonction de degrés de difficulté, mais qui nécessitent une attention et un entraînement particuliers pour avoir une chance réelle de contribuer à la maturité intellectuelle et émotionnelle ainsi qu'à la formation de la personnalité des

⁷⁷ Sous l'appellation de «compétence communicationnelle», qui prête à confusion, on trouve, dans la liste des descripteurs relatifs aux capacités dynamiques des standards de formation autrichiens, des stratégies visant à favoriser la communication, telles que poser des questions de clarification ou être attentif aux signaux non verbaux.

élèves⁷⁸». A l'évidence, il ne s'agit donc plus ici de définition de standards – ni de base ni réguliers – mais de formulation d'objectifs d'apprentissage.

A l'instar des standards autrichiens, les standards de formation allemands (des standards réguliers dans les deux cas) contiennent aussi des descriptions relatives aux compétences interculturelles et méthodologiques. Cependant, ces descriptions ne vont pas sans poser certains problèmes, et ce pour plusieurs raisons: elles ne sont pas justifiées sur le plan théorique par des références à un modèle de compétences; elles ne sont pas ou presque pas formulées en fonction de niveaux; elles ne sont pas illustrées par des exemples d'exercices. Qui plus est, elles mélangent des objectifs de formation tout à fait généraux, applicables dans toutes les disciplines, comme «travailler de manière autonome et savoir coopérer» (KMK) ou «est capable de travailler au sein d'un groupe et d'y partager ses propres idées» (Gassner 2005), avec des objectifs qui sont plus étroitement liés aux langues étrangères, par ex. «est capable de se mettre à la place de son partenaire d'une autre culture en étant ouvert à sa façon de voir et de penser» (KMK) ou «sait que les modes de comportement du quotidien et les principes tels que les règles de circulation, les habitudes à table ou le langage du corps dépendent de la culture de la personne, et qu'ils sont donc fondamentalement équivalents» (Gassner 2005). La sélection (pour les standards) semble parfois quelque peu aléatoire, et il est probable qu'elle n'obtiendrait pas forcément une évaluation convergente de la part de didacticiens spécialisés. Citons par ex.: «répète après le locuteur natif et sait imiter l'intonation et la prononciation du modèle langagier» ou «s'amuse parfois à se faire la conversation dans la langue étrangère» (Gassner 2005).

Lors d'un panel d'experts (membres du conseil consultatif et didacticiens spécialisés invités à la rencontre), nous avons fait évaluer et commenter près de 50 descripteurs relatifs aux compétences interculturelles et méthodologiques, qui provenaient de différentes sources (dont les standards de formation autrichiens et allemands). A cette occasion, il s'est avéré premièrement que, pour plus de la moitié des descripteurs, les priorités fixées étaient passablement différentes; deuxièmement, qu'il ne semblait que rarement possible d'établir un classement clair en fonction des niveaux scolaires de 8^e ou de 11^e année ainsi que de la première ou de la deuxième langues étrangères; et troisièmement, que dans leurs commentaires, les expertes et les experts ont fait, à plusieurs reprises, référence, avec un certain scepticisme, au caractère idéaliste de telles compétences et aux difficultés qui se poseraient si l'on cherchait à les tester.

C'est pour toutes ces raisons que nous avons décidé de ne pas inclure, dans les documents relatifs au modèle de compétences et aux standards de base, des descriptions de dispositions et d'aptitudes interculturelles et méthodologiques, ou encore d'autre ordre, figurant dans le *Cadre de référence* sous les «compétences générales». Nous avons préféré indiquer par un commentaire que les compétences interculturelles et méthodologiques constituent des objectifs de formation centraux de l'école, et qu'elles doivent recevoir une place en conséquence dans les plans d'études régionaux. De plus, nous considérons qu'il vaut mieux, en l'état actuel des connaissances, renoncer à inclure des tests sur ce domaine dans le monitoring de l'éducation et miser entièrement sur l'évaluation (formative) qu'offre le portfolio des langues. En effet, le portfolio des langues permet une évaluation individuelle, réflexive, des expériences tant culturelles et interculturelles que de celles que font les élèves dans leur apprentissage des langues et par rapport à la diversité linguistique. Cette position est aussi confirmée par la littérature sur le sujet, dans laquelle le rôle central de l'auto-évaluation et de l'approche du portfolio est souligné à maintes reprises. Vollmer s'exprime ainsi sur la question: «Pour l'évaluation des *compétences interculturelles*, c'est bien l'auto-évaluation qui me semble de toute façon être la forme première, si ce n'est la seule forme valable, d'évaluation⁷⁹» (cf. par ex. Beacco 2004, 251; Byram / Gribkova / Starkey 2002, 33; Lázár et al. 2007, 29; Scarino 2007, 5).

En exposant notre choix, nous avons évoqué l'état actuel de la recherche, qui sera esquissé de manière synthétique dans les sections suivantes, dans le but de donner des indications en vue de l'élaboration de plans d'études et d'ouvrages d'enseignement, mais aussi dans la perspective des projets qui seront menés dans la lignée d'HarmoS. Les travaux de recherche et de développement dans ces domaines sont en bonne voie. Il est à espérer que des progrès seront faits, par rapport tant à la conceptualisation qu'aux possibilités d'évaluation dans ces domaines. Il faudrait que les

⁷⁸ Trad. libre de deux citations tirées de: Buchberger et al. 2002, 19.

⁷⁹ Trad. libre d'une citation tirée de: Vollmer (2005, 267).

domaines de compétences discutés ici reçoivent une attention particulière dans le cadre des projets qui seront réalisés à la suite d'HarmoS, et que des ressources correspondantes soient aussi libérées pour les enquêtes qui s'avéreront nécessaires. Contrairement à ce qui se fait dans les projets actuels, il faudrait prévoir, de manière ciblée, une collaboration étroite des consortiums dans les différents domaines disciplinaires, notamment une coordination entre les projets conçus pour les langues scolaires (L1) et ceux destinés aux langues étrangères (L2 / L3). Cela signifie qu'il faudrait partir d'un concept de compétence plurilingue et pluriculturelle, et remanier les standards à partir de cette approche, dans la perspective d'une didactique intégrée des langues.

4.2.2 Compétences culturelles et interculturelles

Le domaine des compétences culturelles et interculturelles a été conceptualisé de différentes manières, tant au niveau de ses composantes, de ses dimensions, qu'en ce qui concerne son développement ou son organisation en niveaux.

Composantes / dimensions de la compétence (inter)culturelle

Dans le *Cadre de référence* – qui s'aligne largement sur des travaux préliminaires de Byram et d'autres auteurs – les compétences interculturelles se conçoivent comme faisant partie des compétences générales et sont réparties en quatre catégories de savoir et de savoir-faire:

1. *Savoir*: en font partie, outre les connaissances générales a) le savoir socioculturel sur la culture cible respective, par ex. les traits caractéristiques, spécifiques à la culture, de la vie quotidienne, des relations interpersonnelles, des systèmes de valeurs, des conventions sociales, du langage du corps, notamment, et b) la prise de conscience interculturelle, par ex. connaître les similitudes et les différences, ou encore avoir conscience des stéréotypes existants.

2. *Aptitudes et savoir-faire*: dans cette catégorie, on trouve, outre les aptitudes pratiques, des aptitudes interculturelles, à savoir:

- la capacité d'établir une relation entre la culture d'origine et la culture étrangère;
- la sensibilisation à la notion de culture et la capacité de reconnaître et d'utiliser des stratégies variées pour établir le contact avec des gens d'une autre culture;
- la capacité de jouer le rôle d'intermédiaire culturel entre sa propre culture et la culture étrangère et de gérer efficacement des situations de malentendus et de conflits culturels;
- la capacité à aller au-delà de relations superficielles stéréotypées (Conseil de l'Europe 2001, 82 sqq.).

3. *Savoir-être*: à part les motivations, les styles cognitifs et d'autres facteurs liés à la personnalité, les éléments significatifs ici concernent particulièrement les attitudes des apprenants face à ce qui est étranger et à ce qui est nouveau, notamment:

- ouverture et intérêt envers de nouvelles expériences, les autres, d'autres idées, d'autres peuples, d'autres civilisations;
- volonté de relativiser son point de vue et son système de valeurs culturels;
- volonté et de capacité de prendre ses distances par rapport aux attitudes conventionnelles relatives aux différences culturelles (Conseil de l'Europe 2001, 84).

4. *Savoir-apprendre*: ici, ce sont premièrement des aptitudes générales à l'étude et au travail qui sont mentionnées, comme «coopérer efficacement au travail en groupe et par deux»; deuxièmement, des techniques d'apprentissage et des aptitudes heuristiques, qui se rapportent à l'apprentissage des langues au sens strict, et troisièmement des aptitudes en rapport avec de nouvelles expériences (inter)culturelles:

- la capacité d'apprendre de manière efficace (aux plans linguistique et socioculturel) par l'observation directe et la participation à des actes de communication et par le développement des aptitudes perceptives, analytiques et heuristiques; (...)
- la capacité de l'apprenant à s'accommoder d'une expérience nouvelle (des gens nouveaux, une langue nouvelle, de nouvelles manières de faire, etc.) et de mobiliser ses autres compétences (par exemple par l'observation, l'interprétation de ce qui est observé, l'induction, la mémorisation, etc.) pour la situation d'apprentissage donnée (Conseil de l'Europe 2001, 86).

Comme nous l'avons mentionné, le *Cadre de référence* ne contient aucune assertion relative au développement et à l'organisation en niveaux de ces domaines de compétences. Les

compétences interculturelles ne semblent en principe pas y être rattachées à des compétences linguistiques d'un niveau précis.

Byram (1997) différencie les types de savoir un peu autrement, et nomme, en plus du *savoir* en tant que connaissance proprement dite, un *savoir-comprendre*: une compréhension qui fait intervenir le savoir culturel. Il ajoute un *savoir s'engager*, qui fait appel à des aptitudes d'évaluation critique en rapport avec sa propre culture et celle de l'autre⁸⁰.

De nombreuses autres modélisations de la compétence interculturelle proposent aussi une distinction entre les dimensions du savoir, des attitudes, des aptitudes et de la conscience. Plusieurs des paramètres suivants apparaissent fréquemment dans ce contexte: respect, empathie, souplesse, patience, intérêt, curiosité, ouverture, volonté, sens de l'humour, tolérance à l'ambiguïté, absence de jugement, et d'autres encore (on trouve un bon aperçu des différentes modélisations dans: Humphrey 2007, ainsi que: Sinicrope / Norris / Watanabe 2007). Selon Bolten, «des caractéristiques de la compétence interculturelle régulièrement citées comme étant significatives (l'empathie, la distance au rôle, la tolérance à l'ambiguïté, la capacité d'établir une [méta]communication, etc.) contribuent également à la réussite ou à l'échec de certaines actions qui ne sont pas caractérisées par des situations de chevauchement culturel⁸¹». Dans la perspective d'une promotion de la compétence interculturelle comme tâche transversale de l'école, il rappelle que, par exemple, «l'empathie, la distance au rôle ou encore la souplesse constituent aussi des prérequis intraculturels indispensables», «pour réussir à bien gérer l'action sociale. Dans ce contexte, il paraît logique de comprendre la compétence interculturelle non pas comme un domaine de compétences à part, mais plutôt comme la capacité à combiner le mieux possible des compétences spécifiques, qu'elles soient personnelles, sociales, disciplinaires ou stratégiques, dans des contextes d'action interculturels⁸²».

Le projet européen Leonardo INCA (= *Intercultural Competence Assessment*), qui s'appuie sur le *Cadre de référence* et dans lequel Michael Byram a notamment joué un rôle prépondérant, distingue six éléments de compétences, à savoir: l'empathie, l'ouverture à d'autres cultures, l'acquisition de savoir, la conscience communicationnelle, la tolérance à l'ambiguïté, la capacité d'adapter son comportement à la situation. Ces six éléments sont résumés, pour permettre une meilleure compréhension de l'utilisateur, en trois «piliers» de la compétence interculturelle: *ouverture*, *savoir* et *capacité d'adaptation*:

L'ouverture comprend le respect d'autres cultures et la tolérance à l'ambiguïté.

Le savoir comporte l'acquisition de connaissances et l'empathie.

La capacité d'adaptation se compose de la capacité d'adapter son comportement à la situation et de la conscience communicationnelle. (INCA 2004)

Dans le cadre du projet Socrates LOLIPOP, un portfolio des langues a été conçu en ligne, lequel se propose de mettre un accent particulier sur l'apprentissage interculturel. Il distingue les trois dimensions du savoir, du savoir-faire et du savoir-être, lesquelles sont décrites à l'aide de descripteurs – qui n'ont toutefois pas été vérifiés par voie empirique.

Ces trois types de savoir forment aussi la dimension du concept de compétence interculturelle qui sous-tend le projet du Centre européen pour les langues vivantes de Graz relatif au développement et à l'évaluation d'une compétence de communication interculturelle (Lázár / Huber-Kriegler / Lussier / Matei / Peck 2007).

⁸⁰ Le modèle de Byram (1997) a été utilisé, dans le cadre du projet LACE de l'UE, parallèlement au modèle de Chen / Starosta (1998) et au *Cadre de référence*, pour procéder à une analyse de plans d'études et de la pratique d'enseignement en rapport avec des objectifs interculturels (Franklin et al. 2007). En dépit de certaines lacunes méthodologiques, les résultats de base de l'enquête menée par les pays participants s'avèrent tout à fait probants – même pour la Suisse qui n'y a pas pris part: la valeur et la place des objectifs interculturels est relativement faible dans les plans d'études et l'enseignement actuel. Ce sont les aspects du savoir et de l'attitude qui y trouvent le plus souvent un ancrage; quant à la compétence d'action interculturelle, elle n'y figure par contre qu'assez rarement.

⁸¹ Trad. libre d'une citation de Bolten (2007).

⁸² Trad. libre de deux citations de Bolten (2001).

Dans sa «Proposition pour un référentiel pour les compétences culturelles dans l'enseignement des langues», Jean-Claude Beacco, conseiller de programme de la Division des Politiques linguistiques (DPL) du Conseil de l'Europe adopte une répartition en plusieurs dimensions: savoir, attitudes, situations (d'interaction) et formes de discours. Il se concentre sur les compétences (socio)culturelles et tente de les décrire en restant aussi proche que possible des compétences langagières de communication. Il opère une distinction – pas toujours facile à appréhender – entre les composantes suivantes:

les **dimensions actionnelles** (dites aussi sociopragmatiques) : elles permettent de décrire un «savoir agir» minimum dans une communauté peu connue ou inconnue, de manière à pouvoir y gérer sa vie matérielle et relationnelle (...)

les **dimensions ethnolinguistiques** permettent de caractériser le «vivre ensemble verbal», comme ensemble de normes concernant les comportements communicatifs (relevant de l'ethnologie de la communication) et qui ont des effets directs sur la réussite de la communication. (...)

les **dimensions relationnelles** concernent les attitudes et les compétences verbales nécessaires à une gestion appropriée d'interactions portant explicitement, même partiellement, sur la matière culturelle et sociétale elle-même. Ces interactions ne relèvent pas uniquement d'un savoir-faire d'ordre interactionnel : elles constituent l'une des formes de la «communication verbale interculturelle» et impliquent la constitution ou le renforcement d'attitudes positives comme la curiosité, la bienveillance et la tolérance culturelle ;

les **dimensions interprétatives** concernent la manière dont les utilisateurs / apprenants donnent sens à des sociétés qui ne leur sont pas familières, dont ils ne partagent pas les normes et bien des référents et en rendent compte (pour eux-mêmes et à d'autres). (...)

les **dimensions éducatives ou interculturelles** concernent la matière culturelle en tant qu'elle est objet de jugements évaluatifs, susceptibles, en particulier, d'activer des réactions d'ethnocentrisme et d'intolérance ou des processus d'acculturation et de dépendance culturelle, d'enfermement identitaire ou d'aliénation. (Beacco 2004, 253-254).

Dans le détail, la proposition de Beacco comporte toute une série d'impulsions intéressantes, mais il reste toutefois à un niveau d'abstraction élevé, ce qui ne favorise pas une mise en œuvre dans la pratique de l'élaboration des programmes et des plans d'études.

Le problème des délimitations justifiées sur le plan théorique et en même temps plausibles, et donc le problème des chevauchements et des redondances, qui se pose dans la proposition de Beacco, se fait aussi jour dans le projet CARAP⁸³, et il est d'ailleurs abordé à plusieurs reprises dans le rapport de projet. Dans le cadre de ce projet du Centre européen pour les langues vivantes de Graz, on a tenté de développer un cadre de référence pour des approches plurielles des langues et des cultures. Comme dans le *Cadre de référence*, où l'on retrouve souvent la formule «compétence plurilingue et pluriculturelle», le CARAP cherche aussi à combiner la compétence plurilingue et la compétence interculturelle.

On distingue alors trois «zones»:

1. La zone de la gestion de la communication linguistique et culturelle en contexte d'altérité

Dans ce domaine de compétences relatif à la communication en situation d'hétérogénéité culturelle, on trouve:

- la compétence de résoudre des conflits, de lever des obstacles, de clarifier des malentendus dans le domaine interculturel;
- la compétence de négociation dans l'interaction interculturelle;
- la compétence de médiation linguistique;
- la compétence d'adaptation.

2. La zone de construction et d'élargissement d'un répertoire linguistique et culturel pluriel

Ce domaine de compétences se compose de deux micro-compétences, à savoir:

- la compétence de mettre à profit ses propres expériences interculturelles et plurilingues;
- la compétence d'initier des processus d'apprentissage systématiques et contrôlés, en fonction de contextes interculturels, tant dans le cadre scolaire qu'en dehors.

⁸³ CARAP = Cadre de Référence pour les Approches Plurielles des Langues et des Cultures.

3. Une zone intermédiaire

Cette zone intermédiaire comporte des micro-compétences qui sont considérées comme utiles pour les deux zones susmentionnées, à savoir:

- la compétence de changer de perspective;
- la compétence de donner du sens à l'inconnu linguistique et / ou culturel
- la compétence de prendre du recul;
- la compétence de porter une analyse critique sur sa propre situation de communication ou une situation d'apprentissage et les activités en cause;
- la compétence de reconnaître l'altérité de l'autre dans ses différences et ses similitudes (Candelier et al. 2007).

Si la quantité de descriptions de compétences proposées – principalement tirées de la littérature spécialisée – est impressionnante et si le CARAP peut être mis à profit par les praticiens comme source d'inspiration, on ne peut toutefois pas dire qu'on ait ainsi créé un cadre de référence pour une compétence plurilingue et pluriculturelle. Le cadre ne dépasse guère le stade de la collecte de matériel. C'est la raison pour laquelle une demande de poursuite du projet a été déposée dans le cadre du prochain programme de développement du Centre européen pour les langues vivantes de Graz.

Pour les plans d'études, il faut une répartition des dimensions de la compétence interculturelle plus simple, permettant une bonne vue d'ensemble, facile à saisir pour les auteurs d'ouvrages d'enseignement ainsi que les enseignants, comme c'est le cas dans les standards de formation de la KMK. Cette dernière a en effet opéré une distinction analogue aux trois types de savoir (savoir, savoir-être, savoir-faire) en retenant les domaines suivants:

- savoir d'orientation socioculturel
- gestion compréhensive de la différence culturelle
- maîtrise pratique de situations de rencontre interculturelles (KMK 2005, 8)

La commission souligne aussi à juste titre que le développement des compétences dans ces domaines relève des tâches générales de l'école⁸⁴.

Le développement de compétences interculturelles est une tâche générale qui incombe à l'école, à laquelle l'enseignement des langues étrangères contribue considérablement. Face aux expériences de plus en plus fréquentes de diversité culturelle qui surviennent sur le plan personnel et au travers des médias, il revient aussi à l'enseignement des langues étrangères de faire des élèves des citoyens capables de communiquer et donc ouverts, tolérants et responsables, dans une Europe qui se rapproche. Avec la capacité de comparer, par une approche tolérante et critique, ses propres opinions, valeurs et rapports sociétaux dans certains contextes avec ceux d'autres cultures, et avec la volonté de faire preuve d'intérêt et de compréhension pour les modes de vie et manières de penser, les valeurs, les normes et les conditions de vie de personnes issues d'un autre cercle culturel, les élèves connaîtront un gain d'expérience et vivront un renforcement de leur identité. L'enseignement des langues étrangères contribue surtout à cette perspective plurielle du développement de la personnalité par un savoir d'orientation sur des thèmes et des contenus phares, ainsi que par la transmission de capacités relatives à la communication interculturelle. (KMK 2005, 6)

Les éléments les plus spécifiques pour les disciplines de langues étrangères, ce sont les objectifs et les contenus en rapport avec le savoir socioculturel et culturel. A cet égard, il faut retenir ceci:

- par rapport aux objectifs pour la discipline de la deuxième langue nationale: qu'il y a, toutes régions linguistiques confondues, de nombreux points communs et un savoir socioculturel partagé (en rapport avec les institutions, la politique, l'histoire, etc.);

⁸⁴ Trad. libre d'une citation tirée de la décision de la KMK du 25.10.1996: «Beschuß der Kultusministerkonferenz vom 25.10.1996». In: Bundeszentrale für politische Bildung (éd.), *Interkulturelles Lernen*. Bonn 1998: «L'aspect interculturel ne doit pas alors être cloisonné dans certains thèmes, certaines disciplines ou certains projets; il fait partie en réalité des tâches transversales de l'école.» (trad. libre) Concernant la problématique que pose la mise en œuvre de cette exigence, cf. Bolten 2001.

- qu'on assiste de plus en plus à l'émergence d'une culture jeunes internationale (dans laquelle l'anglais – et la culture américaine – joue un rôle important);
- que la culture dans laquelle vivent les apprenants (école et environnement) ainsi que la culture de la / des langue(s) cible(s) ne sont pas homogènes et fermées sur elles-mêmes, mais qu'elles sont empreintes de mixité et deviennent «transculturelles»; cette «hybridation» de la culture continuera de s'accroître avec l'intensification de la mondialisation.

La compétence d'action interculturelle devrait être encouragée en lien, le plus possible, avec des projets d'échange et de rencontre réels et virtuels. Dans le cadre d'une pédagogie des échanges, il peut aussi faire sens de s'orienter d'après des incidents critiques («critical incidents») possibles, des points névralgiques («hotspots») de la communication interculturelle et des mots clés («hot words») (Heringer 2004). Mais là aussi, il ne faut pas oublier que les situations de rencontre interculturelles ne doivent pas nécessairement être recherchées en dehors de la région linguistique de l'élève; elles peuvent en effet se présenter dans l'environnement des apprenants et dans le cadre d'une classe multiculturelle.

Niveaux de compétence (inter)culturelle

Comme nous l'avons mentionné à plusieurs reprises, le *Cadre de référence* ne contient pas d'échelle relative aux compétences socioculturelles et interculturelles. Il n'avait pas été possible, dans le cadre du programme national de recherche PNR 33, d'étalonner des descripteurs en rapport avec ces domaines de compétences (Schneider / North 2000, 89-90). Différentes tentatives ont été lancées – indépendamment du *Cadre de référence* ou en rapport avec celui-ci – pour établir des niveaux de compétence (inter)culturelle.

Le modèle de développement influent de Milton Bennett, que l'on désigne fréquemment par le sigle DMIS, pour «Developmental Model of Intercultural Sensivity», distingue six niveaux en faisant appel à des aspects cognitifs, affectifs et afférents au comportement⁸⁵. Parmi ces six niveaux, trois sont définis par des orientations ethnocentriques, et les trois autres par des approches «ethnorelatives» (Bennett 1993). La différenciation en six orientations et leur organisation n'ont pu être reproduites empiriquement dans toutes les enquêtes. C'est pourquoi, dans le projet DESI, les six orientations interculturelles du modèle de Bennett ont été «comprises comme des classifications de la compétence interculturelle, sans faire d'hypothèses sur leur organisation⁸⁶». Par ailleurs, les six classes initialement définies ont été synthétisées en trois classes de compétence interculturelle pour des questions de chevauchement et «pour des raisons d'économie de présentation».

La prudence démontrée dans l'enquête DESI fait très clairement ressortir les difficultés qui se posent lorsque l'on veut étayer empiriquement des niveaux de progression pour la compétence interculturelle.

Souvent, c'est avec moins de prudence que des niveaux de compétence interculturelle sont postulés sur une base intuitive. Ainsi, dans certaines versions antérieures du portfolio européen des langues, et récemment aussi dans le portfolio en ligne appelé LOLIPOP, des descriptions de compétences interculturelles des types «savoir», «savoir-comprendre», «savoir-faire», «savoir-être» et «savoir-apprendre» ont été classées de manière intuitive – et de façon assez problématique – en fonction des niveaux de compétence langagière A1-C2 (pour un débat de la problématique, se référer à Schneider / Lenz 2001; Little / Simpson 2003; Lenz / Schneider 2004)⁸⁷.

⁸⁵ Comme le dit le nom du modèle, ce dernier porte sur la sensibilité interculturelle, qui est définie comme la capacité à distinguer et à expérimenter des différences culturelles significatives («ability to discriminate and experience relevant cultural differences»), contrairement à la compétence interculturelle, qui se définit comme la capacité à penser et à agir de manière appropriée dans une situation d'interculturalité («ability to think and act in interculturally appropriate way» - Hammer, Bennet & Wiseman 2003).

⁸⁶ Trad. libre d'une citation tirée de: Hesse / Göbel 2007, 266.

⁸⁷ N'ont par exemple pas non plus été validées empiriquement les propositions de niveaux pour la formation interculturelle de Gogolin (2003) et pour la compétence de traitement des textes littéraires de Burwitz-Melzer (2005 et 2008).

Beacco (2004) classe également ses descriptions de compétences culturelles et interculturelles en fonction des niveaux A1-C2. Sa proposition, qu'il soumet expressément au débat, se veut avant tout une aide de planification pour l'enseignement des langues étrangères. D'ailleurs, lui-même fait remarquer, en toute clairvoyance:

«Ces spécifications des compétences devraient être en mesure de constituer la base d'un référentiel culturel, sans que soient pour autant résolues la question de la diversité des situations éducatives et celle de leur organisation en niveaux successifs de compétences, comme pour les compétences langagières et communicationnelles. C'est peu dire que la détermination de niveaux de maîtrise distincts pour ces compétences culturelles demeure extrêmement problématique» (Beacco 2004, 253).

Le projet INCA établit trois niveaux de compétence: degré 1 = *compétence faible*, degré 2 = *compétence moyenne* et degré 3 = *compétence élevée*. Ces trois degrés sont décrits dans une version pour les utilisateurs (*assessee version*) et une version pour les évaluateurs (*assessor version*). Les deux versions comportent une échelle globale (avec une formulation qui varie légèrement) ainsi que des échelles relatives aux six éléments de compétences que sont: l'empathie, l'ouverture à d'autres cultures, l'acquisition de savoir, la conscience communicationnelle, la tolérance à l'ambiguïté, de même que la capacité d'adapter son comportement à la situation. Il est à noter que, dans la version pour les utilisateurs (*assessee version*), contrairement à celle pour les évaluateurs (*assessor version*), les descriptions de niveaux sont formulées en termes majoritairement positifs. Les descriptions des trois niveaux font référence les unes aux autres – ce ne sont donc pas des formulations isolées, des *stand-alone statements*; elles prennent appui sur le degré de compétence inférieur. Ainsi, la «définition générale» de la compétence moyenne (dans la version pour les utilisateurs) est décrite en ces termes: «Sur la base de mon expérience et de mes acquis, je commence à reconnaître des liens de corrélation dans certains aspects de rencontres interculturelles, que j'ai jusqu'à présent considérés comme étant des événements uniques». La description pour le degré 3 (compétence élevée) commence ainsi: «Maintenant, j'emploie intuitivement un grand nombre des compétences que j'ai développées consciemment au deuxième degré de compétence». La version utilisateurs des échelles à trois degrés est aussi reprise dans le «passeport de la compétence interculturelle» (*Passport of Intercultural Competence*) du portfolio INCA, et doit servir d'«évaluation de performance continue sur le lieu de travail ainsi que dans le cadre de mesures de formation».

De même, trois niveaux différents sont définis – sur la base des travaux de Denise Lussier – dans le projet du Centre européen pour les langues vivantes de Graz qui porte sur le développement et l'évaluation de compétences de communication interculturelles. Lázár et al. se réfèrent, pour l'évaluation sommative, d'une part à l'échelle à trois degrés du projet INCA et, d'autre part – ce qui est peu compréhensible – à l'échelle à six niveaux de Valdes (1986) qui, de par son orientation sur l'idéal du locuteur natif (*native speaker*), ne correspond toutefois guère à l'approche du projet, qui s'inspire sinon du *Cadre de référence* et du projet INCA. Par ailleurs, une échelle à trois niveaux est aussi proposée, sur le CD et sur le site Internet du projet, pour chacun des trois domaines que sont: le savoir interculturel, les aptitudes interculturelles et les attitudes interculturelles. Ces échelles doivent servir d'aide aux évaluateurs et de support pour l'usage en classe. Toutefois, on peut se demander si ces outils sont utilisables car les descripteurs – réunis en sections assez longues, dans le sens d'échelles globales – sont souvent très généraux, vagues, peu uniformes et parfois formulés de manière compliquée. La délimitation des trois domaines ne semble pas toujours claire et plausible. Aucune information précise n'est donnée sur l'origine des descripteurs, ni sur la validation de l'organisation en degrés qui est proposée. A l'évidence, ces descripteurs ne se prêtent pas à la pratique de l'auto-évaluation, et n'ont pas non plus été prévus à cet effet.

Evaluation de compétences interculturelles

Scarino estime quant à lui qu'une évaluation de l'apprentissage interculturel des langues doit satisfaire notamment à une série d'exigences:

Given the features of intercultural language learning, its assessment needs to:

- position the student as both performer and analyser in interaction, though in any individual procedure one or the other role may be foregrounded for different purposes; (...)

- draw upon a range of assessment-types including interviews, conferences, journals, observations, story-telling, and capture students' cumulative learning so that development and progress can be taken into account, for example, through portfolios;
- include self-assessment that recognises learning as a personal process;
- include dimensions that require reflection on the part of students on their developing knowledge and understanding⁸⁸.

Lázár et al. (2007) soulignent eux aussi la nécessité d'inclure une diversité de procédés et d'instruments d'évaluation; parmi les moyens qui permettent d'obtenir des informations utiles, ils citent: «les instruments d'enquêtes en pré- et post-formation ou enseignement, l'auto-évaluation des apprenants, l'observation de la démarche d'apprentissage et des progrès de l'apprenant à l'aide de grilles d'évaluation remplies par l'enseignant, les évaluations de l'enseignant et le portfolio de l'apprenant». Cependant, dans le même document, ils attribuent ensuite, de manière assez simplificatrice, certaines méthodes aux trois domaines du savoir:

«Les méthodes d'évaluation vont donc varier en conséquence afin d'évaluer l'apprenant aussi efficacement que possible: procédures d'évaluation indirectes pour les «savoirs», réalisation de tâches pour le «savoir-faire», auto-évaluation, questionnement sur les attitudes, grilles de l'enseignant et portfolio de l'apprenant pour le «savoir-être». (Lázár et al. 2007)

C'est bien dans le projet INCA que l'évaluation de la compétence interculturelle est mise au point de la manière la plus concrète. L'évaluation INCA comprend trois instruments, ou procédures, désignés comme des «types de tests»:

1. deux questionnaires qui permettent aux évaluateurs d'obtenir des indications sur l'arrière-plan professionnel de la personne évaluée, ainsi que sur son expérience personnelle en situation d'interculturalité;
2. des scénarios: la personne évaluée doit répondre à des questions portant sur des scénarios qui lui sont proposés dans des textes et dans un enregistrement vidéo, et qui présentent des situations interculturelles - chaque question couvrant un aspect de la compétence interculturelle;
3. des jeux de rôle, dans lesquels les personnes évaluées doivent interagir et communiquer selon un rôle imposé. Les jeux de rôle doivent couvrir les six éléments de compétences sans exception.

Différents évaluateurs sont prévus pour les trois types de tests. Les résultats sont discutés lors d'une rencontre des évaluateurs, qui s'entendent sur un résultat global pour les trois piliers de la compétence interculturelle. Cette procédure laborieuse, exécutée à la manière d'un centre d'évaluation, avec des éléments d'évaluation électronique (*e-assessment*), présuppose une formation approfondie des évaluateurs (cf. Bolten 2007). Le concept d'évaluation du projet INCA, axé sur les adultes et leur monde professionnel, ne se prête pas, ne serait-ce que parce qu'il nécessite tant de travail, aux tests qui seront menés dans le cadre du monitoring de l'éducation pour le degré secondaire I.

Caractéristique de ce projet, un «portfolio interculturel» vient compléter la procédure d'évaluation d'INCA. Avec ses trois parties, il fait penser, par bien des aspects, au portfolio européen des langues, et comporte notamment une grille d'auto-évaluation.

⁸⁸ *Compte tenu des caractéristiques de l'apprentissage interculturel des langues, une évaluation dans ce domaine doit:*

- *placer l'étudiant comme sujet agissant et analysant, en interaction, même si, dans toute procédure individuelle, l'un ou l'autre de ces rôles peut être mis en avant pour différentes raisons; (...)*
- *être composée d'une palette de différents types d'évaluation, comme les entretiens, les conférences, les journaux personnels, les observations, les narrations, et saisir l'apprentissage cumulatif des étudiants de façon à prendre en compte leur développement et leurs progrès, par exemple au moyen de portfolios;*
- *inclure un volet d'auto-évaluation, qui reconnaisse l'apprentissage comme étant un processus personnel;*
- *comprendre des dimensions qui nécessitent une réflexion de la part des étudiants quant à l'évolution de leur savoir et de leur compréhension (trad. libre d'une citation de Scarino (2007).*

Comme nous l'avons fait remarquer en introduction, il nous semble approprié de miser, au degré secondaire I, entièrement sur l'évaluation formative dans le travail avec le portfolio des langues. Les indications et les instruments proposés dans le PEL II forment le programme sous une forme compréhensible pour les apprenants (PEL II, Introduction, p. 8):

La biographie langagière t'aidera...

- à *décrire tes expériences de langue et ce que tu as appris sur d'autres personnes, pays et cultures;*
- à *mieux connaître et à respecter les langues et les cultures des personnes qui t'entourent.*

Les élèves apprennent à noter leurs réflexions sur les expériences qu'ils ont faites avec d'autres cultures et d'autres langues, à les discuter, les décrire et les documenter. Parmi les éléments expressément mentionnés, on trouve notamment les expériences acquises au contact de camarades venant d'autres cultures, par la correspondance avec des amis de l'étranger et lors de rencontres extrascolaires.

Au cours du panel d'experts évoqué en introduction, ces derniers ont aussi évalué des descripteurs relatifs à la compétence interculturelle, certains originaux d'autres adaptés, qui avaient été tirés de différentes sources. A cette occasion, l'objectif suivant a été jugé très significatif pour le groupe cible d'HarmoS; il s'agit d'un descripteur qui fait référence au portfolio:

- la volonté et la capacité de consigner, par ex. dans le *portfolio des langues*, des expériences vécues au contact d'autres cultures et ses réflexions à ce sujet.

Les deux autres descripteurs jugés également importants et comportant des objectifs pour l'apprentissage interculturel à l'école étaient les suivants:

- la volonté et la capacité de découvrir le quotidien des autres régions linguistiques (avec l'aide notamment de la langue scolaire locale), de le comparer avec son propre monde d'expérience et de se préparer à rencontrer réellement des locuteurs d'autres régions linguistiques;
- savoir et comprendre qu'il existe plus d'une identité linguistique et plus d'une identité culturelle.

4.2.3 Compétences méthodologiques

Dans les résultats des exercices de tests de communication HarMoS, on découvre des influences qui dénotent la mise en œuvre de différentes stratégies, notamment:

- la répercussion des stratégies d'apprentissage appliquées au fil du parcours scolaire, ainsi que des stratégies de communication favorisant l'apprentissage qui ont été employées dans le passé, dans des situations de communication scolaires et extrascolaires;
- l'effet des stratégies d'utilisation de la langue, qui sont mises en œuvre en situation de test (par ex. des stratégies d'interprétation des textes, des stratégies de planification pour l'écrit ou des stratégies de communication dans l'interaction);
- l'effet des stratégies de test qui sont déployées avant un test ou en situation de test.

L'influence de telles stratégies n'est pas directement saisie ou contrôlée dans les exercices de tests, et aucun standard de base n'a été défini pour les stratégies des apprenants. Cela ne doit en aucun cas être interprété comme une dévalorisation de ces domaines de compétences. Ils doivent au contraire avoir une place centrale dans les plans d'études régionaux, dans l'enseignement ainsi que dans la formation des enseignantes et des enseignants.

Depuis les années quatre-vingt du siècle passé, les stratégies d'apprentissage font l'objet de nombreuses études dans la recherche et la didactique des langues étrangères. On trouve des définitions plus ou moins englobantes et de nombreuses organisations différentes dans ce domaine. Ce qui est caractéristique de cette évolution, c'est que le domaine des stratégies des apprenants a été saisi de façon de plus en plus large et de plus en plus complexe. Ainsi, Rebecca Oxford (1990) a inclus, dans son inventaire de stratégies bien connu (*Strategy Inventory for Language Learning, SILL*), des groupes de stratégies affectives et sociales, ainsi que des

stratégies de mémorisation, en plus des stratégies de communication et des stratégies cognitives ou métacognitives (O'Malley / Chamot 1990).

Récemment, le concept des stratégies des apprenants s'est fortement rapproché des concepts de conscience linguistique, d'éveil aux langues (*language awareness*) d'une part (cf. par ex. Eckerth / Riemer 2000; McIntosh / Noels 2004; Morkötter 2005; Svalberg 2007; Candelier et al. 2007), et d'autonomie des apprenants d'autre part (voir par ex. Wolff 1989; Bimmel / Rampillon 2000). Ainsi, on lit par exemple dans le rapport de la CDIP relatif au projet «JALING»:

«La rencontre avec les langues doit:

- éveiller la conscience et l'intérêt de la personne pour sa (ses) propre(s) langue(s), de même que pour d'autres langues,
- renforcer la motivation à apprendre des langues étrangères et des langues scolaires,
- transmettre une compétence méthodologique dans le sens de stratégies de compréhension, d'apprentissage et d'«étude», et stimuler le travail autonome dans l'apprentissage des langues,
- et aider la personne à développer des capacités métalangagières et métacognitives par la réflexion comparative sur la (les) langue(s) et l'apprentissage des langues⁸⁹.

Si l'on dispose de certains supports pouvant servir d'impulsion pour le développement de tels domaines de compétences (par ex. Balsiger et al. 2005), des activités dans le sens de l'ELBE / EOLE (*éducation et ouverture aux langues à l'école, language awareness, Begegnung mit Sprachen*) ne font toutefois pas encore partie intégrante des plans d'études. De telles activités doivent trouver un ancrage adéquat dans les nouveaux plans d'études régionaux. Cela est également valable pour les aptitudes essentielles de la comparaison, dans le sens d'une didactique intégrée des langues ou d'une didactique du plurilinguisme, ainsi que pour les stratégies d'intercompréhension et de transfert (cf. notamment Berthele 2007; Candelier et al. 2007; Doyé 2006; Edmondson 2001; Meißner 2002, 2004, 2006).

Pendant longtemps, les stratégies des apprenants n'ont été considérées qu'en lien avec une langue étrangère à la fois. Les ouvrages d'enseignement et la pratique proposent surtout des stratégies et des trucs pour la compréhension de l'écrit et l'apprentissage du vocabulaire. Il est assez rare, par exemple, que des stratégies de planification et de remaniement de textes soient explicitement encouragées dans une langue étrangère, et encore plus rare de rencontrer des stratégies d'intercompréhension.

Au chapitre 2 («Kompetenzbereiche für die erste Fremdsprache») des standards de formation élaborés par la KMK pour la première langue étrangère et applicables à la section «Hauptschule» (9^e année), on peut lire, sous le titre «Compétences méthodologiques» (*Methodische Kompetenzen*):

Des compétences méthodologiques sont systématiquement développées dans le cadre des cours de langue étrangère. Les élèves peuvent développer des compétences dans ce domaine. Ils peuvent:

- prendre des décisions concernant des voies d'apprentissage appropriées, en fonction de leur style d'apprentissage personnel,
- utiliser différentes méthodes pour appréhender et analyser des textes oraux ou écrits,
- mettre en œuvre des procédés pour réaliser des textes oraux et écrits,
- appliquer des formes de travail coopératives,
- déployer des techniques et des stratégies d'apprentissage pour étendre leurs connaissances dans la première langue étrangère ainsi que pour acquérir d'autres langues.

Les élèves disposent ainsi des bases qui leur sont nécessaires pour un apprentissage des langues tout au long de la vie⁹⁰.

⁸⁹ Trad. libre d'une citation de: Saudan et al. 2005, 9.

⁹⁰ Trad. libre d'une citation tirée de: KMK 2005, 10; faisant référence à: KMK 2004, 10.

Dans les standards de formation pour la section scolaire moyenne (10^e année), on peut lire la description synthétique suivante (trad. libre d'une citation tirée de: KMK 2004, 10):

«Les élèves disposent ainsi de la capacité à mettre en œuvre un comportement d'apprentissage des langues autonome et coopératif, qui leur servira de fondement pour l'acquisition d'autres langues ainsi que pour un apprentissage des langues autonome tout au long de la vie».

Le chapitre intitulé «Standards für die Kompetenzbereiche» présente ensuite, pour les domaines suivants, des «techniques de travail et des méthodes disciplinaires et transdisciplinaires», que les élèves doivent pouvoir mettre en œuvre:

- Réception de textes (compréhension de l'écrit et de l'oral);
par ex.: «mettre en œuvre différentes techniques d'écoute et de lecture (notamment écoute et lecture globales, sélectives, détaillées) suivant la fonction / tâche à accomplir»;
- Interaction;
par ex.: «respecter les règles fondamentales de la conversation et surmonter des problèmes de compréhension par des questions de clarification et des moyens non verbaux»;
- Production de textes (expression orale et écrite);
par ex.: «appliquer des techniques pour préparer ses propres textes ou présentations, par ex. se noter des mots clés, établir des plans, préparer des rampes d'action, utiliser des images»;
- Stratégies d'apprentissage;
par ex.: «mettre en œuvre des procédés pour mémoriser et retrouver des mots, des moyens de discours»;
- Présentation et utilisation des médias;
«Employer de nouveaux médias pour se procurer de l'information, pour réaliser des interactions de communication et pour apprendre» mais aussi: «Utiliser des techniques de présentation pour exposer les résultats d'un travail (choisir des médias, appliquer des techniques d'articulation et de visualisation)»;
- Apprentissage conscient et organisation de l'apprentissage
par ex.: «travailler de manière autonome et coopérative»; «reconnaître et mettre à profit les conditions d'apprentissage que l'on sait propices, organiser son travail d'apprentissage et gérer son temps»⁹¹.

Si l'on cherche à fixer des standards pour les compétences méthodologiques (des standards réguliers, dans le cas de la KMK), la question se pose de savoir comment il faut procéder pour opérer une sélection justifiée parmi les compétences méthodologiques et pour établir des niveaux de progression.

Les travaux de recherche et les débats menés depuis Wenden / Rubin (1987) au sujet du «bon apprenant en langues» ont montré que le succès de l'apprentissage ne peut pas être attribué à une certaine combinaison de stratégies d'apprentissage. S'il y a des signes indiquant que l'apprenant qui réussit dispose souvent d'une plus grande palette de stratégies, dont il sait bien faire usage (Skehan 1989; Cohen 1998; Dornyei 2005 notamment), la préférence pour certaines stratégies et leur utilisation est fortement déterminée par le type d'apprenant et le style d'apprentissage de la personne. Or les styles d'apprentissage ne se laissent que difficilement modifier (Grotjahn 2003). C'est pourquoi il est essentiel que les apprenants puissent développer le répertoire de stratégies qui est le plus approprié pour eux; cela nécessite donc que l'enseignement en propose une palette diversifiée, et que de nombreuses occasions soient données aux apprenants d'essayer différentes stratégies, de réfléchir aux résultats obtenus avec chacune et de voir lesquelles leur conviennent individuellement (cf. par ex. Tönshoff 1998, 182).

Il s'avère cependant extrêmement difficile de vérifier si des apprenants disposent réellement du répertoire qui leur correspond, et donc s'ils remplissent un standard donné.

Relevant que le *Cadre de référence* ne contient «que des explications générales sans niveaux de progression» pour les compétences méthodologiques, les auteurs des standards de formation allemands ont avancé des «propositions basées sur l'expérience» (KMK 2005, 11). Les descriptions des compétences méthodologiques attendues pour la 9^e année («Hauptschule») et pour la 10^e année (section scolaire moyenne) sont partiellement différentes. Mais il semble s'agir plutôt de différences de formulation qui peuvent s'expliquer par le fait qu'elles ont été élaborées par différents groupes d'auteurs. On n'y trouve guère de véritables différences de contenu, ni un concept clair pour les différents niveaux d'exigences. Parfois, les formulations pour la 9^e année sont par exemple plus générales et semblent de ce fait décrire des compétences plus englobantes que les descriptions détaillées équivalentes pour la 10^e année.

⁹¹ Trad. libre d'un extrait de: KMK 2005, 15-16; KMK 2004, 17-18.

Rampillon (2003) tente d'établir des niveaux d'exigences pour la compétence d'apprentissage et, comme l'indique le titre de sa contribution, de définir un profil minimal pour des «Stratégies d'apprentissage au seuil de L2 / L3». Elle apporte quelques nuances à ce sujet (Rampillon 2003, 90):

Les stratégies d'apprentissage dépendent, dans leur forme et leur nombre, de la créativité et des enseignants et des apprenants. On ne peut donc pas en faire une liste définitive, mais seulement une compilation des méthodes les plus importantes et les plus fréquemment appliquées. C'est dans cette optique qu'a été établie la liste des compétences de base des stratégies d'apprentissage présentée ci-dessous. Elle représente le profil minimal de la compétence d'apprentissage acquise en L2 et fournit aussi, en même temps, la base du transfert de cette compétence dans la L3. On peut à tout moment y faire des compléments.

Rampillon est consciente de ce que sa liste se compose d'hypothèses, et elle soumet expressément sa proposition à débat (Rampillon 2003, 91):

L'une des difficultés de définir ce que devrait être une connaissance de base pour la L2 et / ou la L3 tient au fait qu'il n'existe pas de principes directeurs permettant de décrire un profil minimal ou un profil de développement ou peut-être même un profil de perfection. En outre, le point de départ des connaissances de l'apprenant en matière de stratégies d'apprentissage est hypothétique, puisque l'on ne sait pas non plus ce qu'il lui vient de son apprentissage préscolaire ou de son instruction primaire, ni dans quelle mesure il maîtrise ces connaissances et est capable de s'en servir. Puisque cela relève en partie de la spéculation, il a fallu faire certaines suppositions.

Au sujet de la démarche intéressante qui consiste à différencier des objectifs de stratégies d'apprentissage pour la première et la deuxième langues étrangères, elle fait remarquer que la répartition des différentes stratégies d'apprentissage entre L2 et L3 a «souvent été faite de façon arbitraire, puisqu'on ne donne pas toujours une progression selon les degrés de difficulté. Il s'agit donc souvent d'une distribution plutôt quantitative, où les stratégies d'apprentissage citées sont en fin de compte interchangeables» (Rampillon 2003, 91). Il est à noter que la répartition, pas toujours facile à saisir, des stratégies entre la première et la deuxième langues étrangères devrait être adaptée suivant l'âge qu'ont les apprenants lorsqu'ils commencent à apprendre la deuxième langue étrangère. Rampillon souligne elle-même que «le nombre d'impondérables est encore très élevé à l'heure actuelle et que des recherches supplémentaires s'avèrent nécessaires pour parvenir à des résultats fiables» (Rampillon 2003, 103).

Vollmer (2004) plaide aussi pour une définition de certains niveaux de progression «malgré les difficultés qui se posent et le caractère provisoire des niveaux définis», et ce en rapport avec les domaines de compétences suivants: la compétence de réflexion linguistique⁹² (ou «conscience linguistique» - *Sprachbewusstheit*), la compétence d'apprentissage des langues⁹³ (ou «conscience d'apprentissage des langues» - *Sprachlernbewusstheit*) ainsi que la compétence stratégique⁹⁴ (ou «conscience stratégique» - *Strategiebewusstheit*). Il propose d'intégrer «des mesures concrètes dans un modèle standard de compétences comme dans un modèle de développement didactique. Cet objectif n'est certes pas facile à réaliser, mais malheureusement indispensable à long terme. Seule l'élaboration d'un tel cadre, l'opérationnalisation des compétences et des niveaux de compétences impliqués ainsi que leur examen empirique procurent à une didactique et à une méthodologie du plurilinguisme la perspective qui convient⁹⁵».

A n'en pas douter, il reste encore énormément de travail de recherche à accomplir pour atteindre cet objectif. Il serait souhaitable que des projets faisant suite à HarmoS puissent aussi contribuer à faire avancer ce domaine de recherche. Cependant, le travail qui s'effectue au niveau des plans d'études ne peut attendre que de nouveaux résultats de recherche soient présentés; il doit fixer les objectifs méthodologiques sur la base des connaissances et des expériences dont nous disposons à l'heure actuelle. On trouve d'ailleurs de nombreuses impulsions dans les travaux déjà effectués, dans les diverses listes de compétences méthodologiques à disposition (par ex. Bimmel / Rampillon 2000; Rampillon 2003; KMK 2004, 2005), et surtout dans les nombreux portfolios de langues existants.

⁹² Dans l'original: «Sprachreflexionskompetenz».

⁹³ Dans l'original: «Sprachlernkompetenz».

⁹⁴ Dans l'original: «Strategiekompetenz».

⁹⁵ Trad. libre d'une citation de: Vollmer 2004, 244.

Il faut absolument que des représentants de la langue scolaire, des langues étrangères et des autres disciplines collaborent, tant pour les projets de recherche que pour le travail sur les plans d'études. Car les stratégies d'apprentissage des langues nécessitent de développer la coopération transcurriculaire, comme le fait remarquer le titre de Grenfell / Harris (2004): «a case for cross-curricular collaboration» («un plaidoyer pour la collaboration transcurriculaire - cf. Vollmer 2006a, 2006b). En Suisse aussi, on constate qu'il y a encore des résistances et un certain scepticisme par rapport à la mise en œuvre d'une didactique intégrée des langues. Meißner souligne à juste titre: «Pourtant, au niveau du développement de la compétence d'apprentissage, il s'avérerait justement judicieux d'opter pour une approche relevant d'une didactique du plurilinguisme⁹⁶» (cf. notamment Behr 2007). Mais cela exige nécessairement de la collaboration (Behr 2007, 91; Meißner 2004, 156; Vollmer 2004, 243; Schneider 2007, 150).

Coopérer entre les différentes disciplines implique aussi de travailler avec le portfolio européen des langues qui, en tant qu'instrument non rattaché à une seule discipline, revêt une importance considérable. Le portfolio des langues dans la version pour les enfants et les jeunes de 11 à 15 ans (PEL II) incite à mener une réflexion sur sa propre biographie langagière et sa biographie d'apprentissage linguistique. Il propose, dans ses rubriques «Mes meilleures façons d'apprendre», de nombreuses impulsions incitant l'utilisateur à essayer différentes stratégies, à en prendre conscience, à les appliquer et à les évaluer. Toutefois, pour que le portfolio puisse déployer pleinement ses effets, il ne doit pas seulement être employé ponctuellement et exécuté comme un exercice imposé. De nombreux enseignants considèrent encore que cela représente une charge supplémentaire dans l'horaire que de travailler expressément sur les stratégies et le portfolio des langues. C'est pourquoi il faut encore investir beaucoup dans la formation des enseignantes et des enseignants: ceux-ci devraient avoir l'occasion, dans le cadre de leur formation initiale et continue, d'expérimenter eux-mêmes l'approche d'une didactique intégrée: dans des cours de langues, lors d'événements et de projets faisant intervenir plusieurs langues, durant leurs stages et au travers de divers échanges. Dans le cadre de leur formation initiale et continue, ils devraient non seulement se familiariser avec le portfolio des langues et apprendre à l'employer, dans toutes ses versions, en fonction des différents groupes d'apprenants, mais ils devraient aussi avoir à remplir leur propre portfolio avec des documents, des comptes rendus et des réflexions personnelles sur leur propre biographie langagière, leurs parcours et leurs habitudes d'apprentissage, de même que leurs stratégies en la matière (Schneider 2007, 150).

⁹⁶ Trad. libre d'une citation de: Meißner 2006, 163.

4.2.4 *Références bibliographiques relatives aux compétences transversales*

Compétences interculturelles

- Bättig, Brigitte / Ruf, Barbara (2005): HARMOS: Lehrplanuntersuchung erste und zweite Fremdsprachen. EDK: Informationsstelle IDES.
- Beacco, Jean-Claude (2004): Une proposition de référentiel pour les compétences culturelles dans les enseignements de langues. In : Beacco, Jean-Claude/ Bouquet, Simon/ Porquier, Rémy (2004): Niveau B2 pour le Français. Textes et Références. Paris: Didier, 251-286.
- Bennett, Milton .J. (1993). Towards ethnorelativism: A developmental model of intercultural sensitivity. In: M. Paige (Ed.), Education for the intercultural experience. Yarmouth, ME: Intercultural Press, 21-71.
- Bersinger, Simone / Jordi, Urs / Tschang, Martine (2005): Europäisches Sprachenportfolio - European Language Portfolio - Portfolio Européen des langues - Version für Kinder und Jugendliche von 11 bis 15 Jahren ESP II. Bern: Schulverlag blmv.
- Boeckmann, Klaus-Börge (2006): Dimensionen von Interkulturalität im Kontext des Fremd- und Zweitsprachenunterrichts. In: Zeitschrift für Interkulturellen Fremdsprachenunterricht [Online] 11/3, 14 S. Online: <http://www.ualberta.ca/~german/ejournal/Boeckmann1.htm> (23.01.08).
- Bolten, Jürgen (2007): Was heißt „Interkulturelle Kompetenz?“ Perspektiven für die internationale Personalentwicklung. In: V.Künzer/ J. Berninghausen (Hrsg.): Wirtschaft als interkulturelle Herausforderung. Frankfurt/M, 21-42. Online: http://www2.uni-jena.de/philosophie/iwk/publikationen/interkulturelle_kompetenz_personalentwicklung_bolten.pdf (05.05.08).
- Bolten, Jürgen (2001):Thesen zum interkulturellen Lernen in der Schule. In: Bolten, Jürgen / Schröter, Daniela (Hrsg.): Interkulturelle Kommunikation. Sternenfels/ Berlin: Verlag Wissen und Praxis.
- Buchberger, Gerlinde / Horak, Angela / Matzer, Edith / Mewald, Claudia / Seidlhofer Barbara / Spenger, Jörg / Stefan, Ferdinand (2002). Standards für die 8. Schulstufe Fachbereich Fremdsprachen (am Beispiel des Englischen). Arbeitsentwürfe für die Hand der Lehrer/innen. Erprobungsversion, Teil II. Online: [http://arge.stvg.at/arge.nsf/073757dce34766ffc12569960053dced/00b8deacb440a538c1256c5b002ae95d/\\$FILE/Standards_Erprobungsversion_Teil_2_Juni_02.pdf](http://arge.stvg.at/arge.nsf/073757dce34766ffc12569960053dced/00b8deacb440a538c1256c5b002ae95d/$FILE/Standards_Erprobungsversion_Teil_2_Juni_02.pdf). (HTML-Version 24.01.08).
- Burwitz-Melzer, Eva (2008): Interkulturelles und sprachliches Lernen mit fremdsprachlichen literarischen Texten: Zwei zentrale Elemente eines neuen Lesekompetenzmodells. In: FLUL 35, 104-120.
- Burwitz-Melzer, Eva (2005): Kompetenzen für den Literaturunterricht heute. Ein Beitrag zur standardorientierten Didaktik des Fremdsprachenunterrichts . In: FLUL 34, 94-120.
- Byram, Michael (2003): Intercultural competence. Strasbourg: Council of Europe.
- Byram, Michael / Gribkova, Bella / Starkey, Hugh (2002): Développer la dimension interculturelle de l'enseignement des langues. Une introduction pratique a l'usage des enseignants. Strasbourg: Conseil de l'Europe.
- Byram, Michael / Neuner, Gerhard / Zarate, Geneviève (1997): Sociocultural competence in language learning and teaching. Studies towards a Common European Framework of reference for language teaching and learning. Strasbourg: Council of Europe Publishing.
- Candelier, Michel / Camilleri-Grima, Antoinette / Castellotti, Véronique / de Pietro, Jean-François / Lörincz, Ildikó / Meissner, Franz-Joseph / Schröder-Sura, Anna / Noguero, Artur (2007): CARAP Cadre de Référence pour les Approches Plurielles des Langues et des Cultures. Version 1. Graz: Centre Européen pour les Langues Vivantes.
- Candelier, Michel / Camilleri-Grima, Antoinette / Castellotti, Véronique / de Pietro, Jean-François / Lörincz, Ildikó / Meissner, Franz-Joseph / Schröder-Sura, Anna / Noguero, Artur (2007): CARAP - Framework of reference for pluralistic approaches to languages and cultures . Version 2. Graz: Centre Européen pour les Langues Vivantes.
- Chen, Guo-Ming/ Starosta, William G. (1998): Foundations of Intercultural Communication. Needham Height, Mass: Allyn and Bacon.
- Condat, Sophie (2008): L'éducation interculturelle. Bibliographie. Paris: Centre de ressources documentaires du CIEP. Online: http://www.ciep.fr/bibliographie/Education_interculturelle.pdf (24.04.08).
- Coste, Daniel (2007): Contextualiser les utilisations du Cadre européen commun de référence pour les langues. In : Conseil de l'Europe : « Le Cadre européen commun de référence pour les langues (CECR) et l'élaboration de politiques linguistiques : défis et responsabilités ». Forum intergouvernemental sur les politiques linguistiques. Rapport, 44-53. Online: http://www.coe.int/t/dg4/linguistic/Source/SourceForum07/ForumFeb07_%20Report_0807_FR.doc (16.07.08).
- Doyé, Peter (2006): Interkomprehension. In: ZFF - Zeitschrift für Fremdsprachenforschung 17 / 2, 245-256.
- EU, European Commission Directorate General for Education and Culture, Multilingualism Policies: Languages and cultures in Europe. LACE. Website: <http://www.lace2007.eu/> (17.07.08).
- Franklin, Peter / Langé, Gisella / Marsh, David / Spencer-Oatey, Helen (2007): Studie über die Vermittlung interkultureller Kompetenzen im Fremdsprachenunterricht im Pflichtschulbereich in der Europäischen Union. Sprachen und Kulturen in Europa (LACE). Online: <http://www.lace2007.eu/docs/de/final.pdf> (06.05.08).

- Gassner, Otmar/ Horak, Angela/ Mewald, Claudia/ Moser, Wolfgang/ Schober, Michael/ Stefan, Ferdinand/ Valsky, Claudia (2005): Bildungsstandards in Österreich. Fremdsprachen. Englisch 8. Schulstufe. Version September'05. Wien: Bundesministerium für Bildung, Wissenschaft und Kultur.
- Glaboniat, Manuela / Müller, Martin / Rusch, Paul / Schmitz, Helen / Wertenschlag, Lukas (2005): Profile deutsch. Gemeinsamer europäischer Referenzrahmen. Lernzielbestimmungen - Kannbeschreibungen - Kommunikative Mittel - Niveau A1 - A2 - B1 - B2 - C1 - C2 mit CD-ROM. München: Langenscheidt.
- Gogolin, Ingrid (2003): Fähigkeitsstufen der Interkulturellen Bildung. Hamburg: Institut für International und Interkulturell Vergleichende Erziehungswissenschaft. Mimeo. Online: http://www.erzwiss.uni-hamburg.de/Personal/Gogolin/cosmea/core/corebase/mediabase/foermig/website_gogolin/dokumente/Gogolin_2002_F_higkeitsstufen_Mimeo.pdf (17.07.08).
- Hammer, M. R., Bennett, M. J., & Wiseman, R. (2003): Measuring intercultural sensitivity: The intercultural development inventory. In: *International Journal of Intercultural Relations*, 27, 421–443.
- Heringer, Hans Jürgen (2004): *Interkulturelle Kommunikation*. Tübingen: Francke (=UTB 2550).
- Hesse, Hermann-Günter / Göbel, Kerstin (2007): *Interkulturelle Kompetenz*. In: Beck, Bärbel/ Klieme, Eckhard (Hrsg.): *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International)*. Weinheim und Basel: Beltz, 256-272.
- Humphrey, Donna (2007): *Intercultural Communication Competence: The State of Knowledge*. Report prepared for CILT The National Centre for Languages : Development of NOS in intercultural Skills Project. Online: <http://www.cilt.org.uk/standards/donnareport.pdf>
- INCA (2004.): *Intercultural Competence Assessment*. Online: <http://www.incaproject.org/> und http://www.incaproject.org/de_downloads/21_INCA_Assessor_Manual_germ.pdf (21.01.08).
- KMK, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2004) (Hrsg.): *Bildungsstandards für die erste Fremdsprache (Englisch/ Französisch) für den Mittleren Schulabschluss*. Neuwied: Luchterhand.
- KMK, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2005) (Hrsg.): *Bildungsstandards für die erste Fremdsprache (Englisch/ Französisch) für den Hauptschulabschluss*. Neuwied: Luchterhand.
- Lázár, Ildikó (2003): *Incorporating intercultural communicative competence in language teacher education*. Strasbourg: Conseil de l' Europe.
- Lázár, Ildikó (2007): *Developing and assessing intercultural communicative competence*. In: *Babylonia* 3, 9-13.
- Lázár, Ildikó / Huber-Kriegler, Martina / Lussier, Denise / Matei, Gabriela S. / Peck, Christiane (2008): *Développer et évaluer la compétence en communication interculturelle - Un guide à l'usage des enseignants de langues et des formateurs d'enseignants*. Graz : Centre européen pour les langues vivantes.
- Lázár, Ildikó / Huber-Kriegler, Martina / Lussier, Denise / Matei, Gabriela S. / Peck, Christiane (Hrsg.) (2007): *Developing and assessing intercultural communicative competence – A guide for language teachers and teacher educators*. Graz: European Centre for Modern Languages.
- Lazar, Ildiko / Huber-Kriegler, Martina / Lussier, Denise / Matei, Gabriela S. / Peck, Christiane (2008): *Développer et évaluer la compétence en communication interculturelle - Un guide à l'usage des enseignants de langues et des formateurs d'enseignants avec cdr*. Strasbourg: Council of Europe.
- Lenz, Peter / Studer, Thomas (2005): *Neue Instrumente für die Beurteilung der Französisch- und Englischkompetenzen von Deutschschweizer Schülerinnen und Schülern*. In: Aline Gohard-Radenkovic (Hrsg.), *Mehrsprachigkeit, Interkulturalität und Fremdsprachendidaktik in einem zweisprachigen Kontext, Transversales*, Bern: Peter Lang, 37-56.
- Little, David (2006): *State-of-the-Art Article: The Common European Framework of Reference for Languages: Content, purpose, origin, reception and impact*. In: *Language Teaching* 2006, 3.
- Little, David / Simpson, Barbara (2003): *Portfolio européen des langues. La composante interculturelle et Apprendre à apprendre*. Strasbourg: Conseil de l'Europe. Online: <http://www.coe.int/T/DG4/Portfolio/documents/Maquettes.pdf> (24.01.08).
- Little, David / Hodel, Hans-Peter / Kohonen, Viljo / Meijer, Dick / Perclová, Radka (2008): *Préparer les enseignants du Portfolio européen des langues - arguments, matériels et ressources*. Livre et CDROM. Graz: European Centre for Modern Languages.
- LOLIPOP: *Language On-line Portfolio Project*. <http://www.isoc.siu.no/isocii.nsf/projectlist/116998> (17.07.08).
- Newby, David (2007): *The European Portfolio for Student Teachers of Languages*. In: *Babylonia*. 3, 23-26.
- Newby, David / Allan, Rebecca / Fenner, Anne-Brit / Jones, Barry / Komorowska, Hanna / Soghikyan, Kristine (2008): *Portfolio européen pour les enseignants en langues en formation initiale - Un outil de réflexion pour la formation des enseignants en langues*. Strasbourg: Conseil de l'Europe.
- Scarino, Angela (2007): *Assessing intercultural language learning*. Australian Government Quality Teacher Program: *Intercultural Language Teaching and Learning in Practice*. Discussion Paper 6. Online: http://www.iltip.unisa.edu.au/doclibpapers/iltip_paper6.doc (23.01.08).

- Schneider, Günther (2003): Das Europäische Sprachenportfolio - Grundidee, Modelle, Erfahrungen. In: Schneider, Günther / Clalüna, Monika (Hrsg.): Mehr Sprache - mehrsprachig - mit Deutsch. Didaktische und politische Perspektiven, 256-275.
- Schneider, Günther (2005): Der ""Gemeinsame europäische Referenzrahmen für Sprachen"" als Grundlage von Bildungsstandards für die Fremdsprachen - Methodologische Probleme der Entwicklung und Adaptierung von Kompetenzbeschreibungen. In: Schweizerische Zeitschrift für Bildungswissenschaften 1, 13-37.
- Sercu, Lies (2002): Autonomes Lernen im interkulturellen Fremdsprachenunterricht. Kriterien für die Auswahl von Lerninhalten und Lernaufgaben. Zeitschrift für Interkulturellen Fremdsprachenunterricht [Online], 7/2, 16 S. Online: <http://zif.spz.tu-darmstadt.de/jg-07-2/beitrag/sercu1.htm> (23.01.08).
- Sercu, Lies (2004): Assessing intercultural competence: a framework for systematic test development in foreign language education and beyond. In: Intercultural Education 15/1, 73-90.
- Sinicrope, Castle., Norris, John .M., and Watanabe, Yukiko. (2007): Understanding and assessing intercultural competence: A summary of theory, research, and practice. Technical report for thForeign Language Program Evaluation Project. Honolulu: University of Hawai'i, National Foreign Language Resource Center. Online: <http://nflrc.hawaii.edu/evaluation/files/ICC%20Final%20Report%20for%20publication.doc> (25.01.08).
- Vollmer, Hermut J. (2005): Das Gespenst der Standardisierung geht um ODER: Lehren und Lernen fremder Sprachen auf der Basis von Bildungsstandards. In: Bausch, Karl-Richard/ Burwitz-Melzer, Eva/ Königs, Frank G. (Hrsg.): Bildungsstandards für den Fremdsprachenunterricht auf dem Prüfstand. Arbeitspapiere der 25. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr, 258-271.
- Zarate, Geneviève / Gohard-Radenkovic, Aline / Lussier, Denise / Penz, Hermine (2003). Médiation culturelle et didactique des langues. Strasbourg: Centre européen pour les langues vivantes. Graz/Strasbourg: Editions du Conseil de l'Europe.

Compétences méthodologiques

- Babylonia. Themenheft: Begegnung mit Sprachen / Ouverture aux langues. 1999/2.
- Babylonia. Themenheft: Strategien für den Sprachunterricht und das Sprachenlernen. 2002/2.
- Babylonia. Themenschwerpunkt: Integrierte Sprachendidaktik. 2005/4.
- Balsiger, Claudine / Berger, Claudia / Dufour, Janine / Gremion, Lise / de Pietro, Danièle / Zurbriggen, Elisabeth (2003): Education et ouverture aux langues à l' école eole. Volume 2 3e année primaire - 6e année. Documents reproductibles. Neuchâtel: SG / CIIP.
- Bausch, Karl-Richard / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.) (2004): Mehrsprachigkeit im Fokus. Arbeitspapiere der 24. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr (= Giessener Beiträge zur Fremdsprachendidaktik).
- Bausch, Karl-Richard / Burwitz-Melzer, Eva / Königs, Frank G. (2005): Bildungsstandards für den Fremdsprachenunterricht auf dem Prüfstand. Arbeitspapiere der 25. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr (= Giessener Beiträge zur Fremdsprachendidaktik).
- Bausch, Karl-Richard / Burwitz-Melzer, Eva / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.) (2006): Aufgabenorientierung als Aufgabe. Arbeitspapiere der 26. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr (= Giessener Beiträge zur Fremdsprachendidaktik).
- Behr, Ursula (2007): Sprachenübergreifendes Lernen und Lehren in der Sekundarstufe I. Ergebnisse eines Kooperationsprojektes der drei Phasen der Lehrerbildung. Tübingen: Narr (= Giessener Beiträge zur Fremdsprachendidaktik).
- Bernaus ,Mercè Andrade, Ana Isabel / Kervran, Martine / Murkowska, Anna / Trujillo Sáez, Fernando (2007) : La dimension plurilingue et pluriculturelle dans la formation des enseignants de langues : kit de formation. Strasbourg / Graz: Conseil de l'Europe / Centre européen pour les langues vivantes. Online : http://www.ecml.at/mtp2/Lea/html/LEA_E_Results.htm (05.05.08).
- Bersinger, Simone / Jordi, Urs / Tschang, Martine (2005): Europäisches Sprachenportfolio - European Language Portfolio - Portfolio Européen des langues - Version für Kinder und Jugendliche von 11 bis 15 Jahren ESP II. Bern: Schulverlag blmv.
- Berthele, Raphael (2007): Zum Prozess des Verstehens und Erschliessens In: Hufeisen, Britta / Marx, Nicole (Hrsg.) (2007): EuroComGerm - Die sieben Siebe: Germanische Sprachen lesen lernen. Aachen: Shaker, 15-26.
- Bimmel, Peter / Rampillon, Ute (2000): Lernerautonomie und Lernstrategien. München: Langenscheidt.
- Candelier, Michel (2003): L' éveil aux langues à l' école primaire. Evlang: bilan d'une innovation européenne. Bruxelles: De Boeck (= Pratiques pédagogiques).
- Candelier, Michel et al. (2003): Janua Linguarum. La porte des langues L'introduction de l'éveil aux langues dans le curriculum. Strasbourg/Graz : Centre européen pour les langues vivantes / Editions du Conseil de l'Europe.
- Candelier, Michel / Camilleri-Grima, Antoinette / Castellotti, Véronique / de Pietro, Jean-François / Lörincz, Ildikó / Meissner, Franz-Joseph / Schröder-Sura, Anna / Noguero, Artur (2007): CARAP - Framework of reference for pluralistic approaches to languages and cultures . Version 2. Graz: Centre Européen pour les Langues Vivantes.

- Chamot, Anna Uhl (2004): Issues in Language Learning Strategy Research and Teaching. In: Electronic Journal of Foreign Language Teaching 1/1, 14-26. Online: <http://e-flt.nus.edu.sg/v1n12004/chamot.htm> (23.01.08).
- Cohen, Andrew (1998): Strategies in learning and using a second language. London: Longman.
- Department for children, schools and families (o.J.): Key Stage 2 Framework for languages. Language learning strategies (LLS): Planning: Analysing and evaluating ways of learning. Online: <http://www.standards.dfes.gov.uk/primary/publications/languages/framework/lfs/1205069/#1205935> (25.01.08).
- Dornyei, Zoltan (2001): Motivation strategies in the language classroom. Cambridge: University Press (= Cambridge Language Teaching Library).
- Dornyei, Zoltán (2005): The Psychology of the Language Learner. Individual Differences in Second Language Acquisition. London: Erlbaum.
- Doyé, Peter (2006): Interkomprehension. In: ZFF - Zeitschrift für Fremdsprachenforschung 17 / 2, 245-256.
- Eckerth, Johannes / Riemer, Claudia (2000): Awareness und Motivation: Noticing als Bindeglied zwischen kognitiven und affektiven Faktoren des Fremdsprachenlernens. In: Riemer, Claudia (Hrsg.): Kognitive Aspekte des Lehrens und Lernens von Fremdsprachen. Festschrift für Willis J. Edmondson zum 60. Geburtstag, 228-246.
- Edmondson, Willis J. (2001): Transfer beim Erlernen einer weiteren Fremdsprache: die L1-Transfer-Vermeidungsstrategie. In: Aguado, Karin / Riemer, Claudia: Wege und Ziele. Zur Theorie, Empirie und Praxis des Deutschen als Fremdsprache (und anderer Fremdsprachen). Festschrift für Gerd Henrici zum 60. Geburtstag. Hohengehren: Schneider, 137-154.
- Edmondson, Willis / House, Juliane (2006, 3. Aufl.): Einführung in die Sprachlehrforschung. Tübingen: Francke (= UTB 1697).
- Faerch, Claus / Kasper, Gabriele (1983): Strategies in Interlanguage Communication. London u. a.: Longman.
- Glaboniat, Manuela / Müller, Martin / Rusch, Paul / Schmitz, Helen / Wertenschlag, Lukas (2005): Profile deutsch. Gemeinsamer europäischer Referenzrahmen. Lernzielbestimmungen - Kannbeschreibungen - Kommunikative Mittel - Niveau A1 - A2 - B1 - B2 - C1 - C2 mit CD-ROM. München: Langenscheidt.
- Grenfell, Michael / Harris, Vee (2004): Language-learning strategies: a case for cross-curricular collaboration. In: Language Awareness, 13/2, 116-130. Online: <http://www.multilingual-matters.net/la/013/0116/la0130116.pdf> (17.07.08).
- Grotjahn, Rüdiger (2003): Lernstile/Lerntypen. In: Bausch, Karl-Richard / Christ, Herbert / Krumm, Hans-Jürgen (Hrsg.): Handbuch Fremdsprachenunterricht. Tübingen: Francke, 326-331.
- Harris, Vee (2001): Helping learners learn: exploring strategy instruction in language classrooms across Europe. Strasbourg: European Centre for Modern Languages / Conseil de l'Europe.
- Holec, Henri / Little, David / Richterich, René (1996): Strategies in language learning and use. Studies towards a Common European Framework of reference for language learning and teaching. Strasbourg: Council of Europe Publishing.
- Hufeisen, Britta / Marx, Nicole (Hrsg.) (2007): EuroComGerm - Die sieben Siebe: Germanische Sprachen lesen lernen. Aachen: Shaker.
- Jessner, Ulrike (2006): Linguistic Awareness in Multilinguals. English as a Third Language. Edinburgh: University Press.
- Klein, Horst G. / Rutke, Dorothea (Hrsg.) (2004): Neuere Forschungen zur Europäischen Interkomprehension. Aachen: Shaker (= Editiones EuroCom 21).
- Kleppin, Karin (2006): Selbstreflexion und Selbstevaluation : ein vernachlässigtes Potential bei Aufgaben. In: Bausch, Karl-Richard / Burwitz-Melzer, Eva / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.): Aufgabenorientierung als Aufgabe. Arbeitspapiere der 26. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr, S. 102-108.
- Kohonen, Viljo (2000): Student reflection in portfolio assessment: making language learning more visible. In: Babylonia. Themenheft: Lernerfahrungen reflektieren - Lernerfolge beurteilen, 1, 15-18.
- Little, David / Hodel, Hans-Peter / Kohonen, Viljo / Meijer, Dick / Perclová, Radka (2008): Préparer les enseignants du Portfolio européen des langues - arguments, matériels et ressources. Livre et CDROM. Strasbourg: Conseil de l'Europe.
- Macaro, Ernesto (2001): Learning strategies in foreign and second language classrooms. London: Continuum.
- McIntosh, Cameron N. / Noels, Kimberly A. (2004). Self-Determined Motivation for Language Learning: The Role of Need for Cognition and Language Learning Strategies. Zeitschrift für Interkulturellen Fremdsprachenunterricht, Online] 9(2), Erhältlich unter http://zif.spz.tu-darmstadt.de/projekt_ejournal/jg-09-2/beitrag/Mcintosh2.htm (30.05.08).
- Meißner, Franz-Joseph (2002): Transfer aus der Sicht der Mehrsprachigkeitsdidaktik. In: Materialien Deutsch als Fremdsprache 65, 128-142.
- Meißner, Franz-Joseph (2004): Sprachenlernen zusammendenken. Aufhören, Rücken an Rücken zu forschen, zu planen, zu lehren und zu lernen. In: Bausch, Karl-Richard / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.) (2004): Mehrsprachigkeit im Fokus. Arbeitspapiere der 24. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr, 148-162.

- Meißner, Franz-Joseph (2006): Lernerautonomisierung durch Aufgaben? In: Bausch, Karl-Richard / Burwitz-Melzer, Eva / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.): Aufgabenorientierung als Aufgabe. Arbeitspapiere der 26. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr, S. 159-170.
- Meißner, Franz-Joseph / Senger, Ulrike (2001): Vom induktiven zum konstruktiven Lehr- und Lernparadigma. Methodische Folgerungen aus der mehrsprachigkeitsdidaktischen Forschung. In: Meißner, Franz-Joseph / Reinfried, Marcus (Hrsg.): Bausteine für einen neokommunikativen Französischunterricht: Lernerzentrierung, Ganzheitlichkeit, Handlungsorientierung, Interkulturalität, Mehrsprachigkeitsdidaktik. Tübingen: Narr, S. 22-51.
- Morkötter, Steffi (2005): Language Awareness und Mehrsprachigkeit. Frankfurt am Main u. a.: Lang (= Kolloquium Fremdsprachenunterricht 21).
- Newby, David (2007): The European Portfolio for Student Teachers of Languages. In: Babylonia. Themenheft: Projekte des Europäischen Fremdsprachenzentrums in Graz. 3, 23-26.
- Newby, David / Allan, Rebecca / Fenner, Anne-Brit / Jones, Barry / Komorowska, Hanna / Soghikyan, Kristine (2008): Portfolio européen pour les enseignants en langues en formation initiale - Un outil de réflexion pour la formation des enseignants en langues. Strasbourg: Conseil de l'Europe.
- O'Malley, Michael J. / Chamot, Anna Uhl (1990): Learning Strategies in Second Language Acquisition. Cambridge: Cambridge University Press (= Cambridge Applied Linguistics).
- Oxford, Rebecca (1990): Language Learning Strategies: What Every Teacher Should Know. New York: Newbury House.
- Rampillon, Ute (2003): Lernstrategisches Minimalprofil an der Schwelle von L2 und L3. In: Hufeisen, Britta / Neuner, Gerhard (Hrsg.): Mehrsprachigkeitskonzept - Tertiärsprachenlernen - Deutsch nach Englisch, 87-106.
- Rampillon, Ute / Zimmermann, Günther (Hrsg.) (1997): Strategien und Techniken beim Erwerb fremder Sprachen. Ismaning: Hueber (= Forum Sprache).
- Robinson, Peter (ed.) (2002): Individual Differences and Instructed Language Learning. Amsterdam: Benjamins (= Language Learning & Language Teaching 2).
- Saudan, Victor (2007): Ursache Zukunft: Die Neukonzeption des Fremdsprachenunterrichts und ihre Auswirkungen auf die Lehrerinnen- und Lehrerbildung. In: Beiträge zur Lehrerbildung 25/2, 214-222.
- Saudan, Victor / Perregaux, Christiane / Mettler, Monika / Deschoux, Carole-Anne / Sauer, Esther / Ladner, Esther (2005): Lernen durch die Sprachenvielfalt. Schlussbericht zum Projekt JALING Suisse. Apprendre par et pour la diversité linguistique. Rapport final sur le projet JALING Suisse. Bern: EDK. Online: <http://www.edk.ch/dyn/17206.php> (20.07.08).
- Schneider, Günther (2007): Fremdsprachenforschung und die Ausbildung von Fremdsprachenlehrerinnen und -lehrern. In: Beiträge zur Lehrerbildung 25/2, 143-155.
- Skehan, Peter (1989): Individual Differences in Second-Language Learning. London: Arnold.
- Steinhauer, Britta (2006): Transfer im Fremdspracherwerb. Ein Forschungsüberblick und eine empirische Untersuchung des individuellen Transferverhaltens. Frankfurt am Main: Lang (= Europäische Hochschulschriften: Reihe 21, Linguistik 294).
- Svalberg, Agneta M-L. (2007): Language awareness and language learning. In: Language teaching. Surveys and Studies October 4 / 40, 287-308.
- Tönshoff, Wolfgang (1997): Bewusstmachendes Training von Fremdsprachenstrategien. In: Materialien Deutsch als Fremdsprache 46, 96-111.
- Tönshoff, Wolfgang (1998): Zur Rolle des Begriffs ‚Kognition‘ beim Lernen(lernen) von Fremdsprachen. In: Bausch, Karl-Richard / Christ, Herbert / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.) Kognition als Schlüsselbegriff bei der Erforschung des Lehrens und Lernens fremder Sprachen, 179-186.
- Vollmer, Helmut J. (1996): Zum Stand der Erforschung des Fremdspracherwerbs in der Bundesrepublik Deutschland - Fokus: Der Lerner. In: Bausch, Karl-Richard / Christ, Herbert / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.): Erforschung des Lehrens und Lernens fremder Sprachen. Zwischenbilanz und Perspektiven. Arbeitspapiere der 16. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr, 144-158.
- Vollmer, Helmut J. (1998): Dem Lerner auf der Spur: Kognitive, emotionale und interaktive Aspekte des Fremdspracherwerbs. In: Bausch, Karl-Richard / Christ, Herbert / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.) Kognition als Schlüsselbegriff bei der Erforschung des Lehrens und Lernens fremder Sprachen, 197-206.
- Vollmer, Helmut J. et al. (2001): Lernen und Lehren von Fremdsprachen: Kognition, Affektion, Interaktion. Ein Forschungsüberblick. In: Zeitschrift für Fremdsprachenforschung 2/2001, 1-143.
- Vollmer, Helmut (2004): Auf dem Wege zu Mehrsprachigkeit – Ansätze, Erfahrungen, Aufbruch. In: Bausch, Karl-Richard / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.) (2004): Mehrsprachigkeit im Fokus. Arbeitspapiere der 24. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr, 238-248.
- Vollmer, Helmut (2005): Das Gespenst der Standardisierung geht um ODER: Lehren Lernen fremder Sprachen auf der Basis von Bildungsstandards. In: Bausch, Karl-Richard / Burwitz-Melzer, Eva / Königs, Frank G. (2005): Bildungsstandards für den Fremdsprachenunterricht auf dem Prüfstand. Arbeitspapiere der 25. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr, 258-271.

- Vollmer, Helmut (2006a): Towards a Common European Instrument for Language(s) of Education. Preliminary Study. Strasbourg: Language Policy Division. Online: http://www.coe.int/t/dg4/linguistic/Source/Vollmer_Beacco_final_EN.doc (20.07.08).
- Vollmer, Helmut (2006b): Language Across the Curriculum. Preliminary Study. Strasbourg: Language Policy Division. Online: http://www.coe.int/t/dg4/linguistic/Source/Vollmer_LAC_EN.doc (20.07.08).
- Vollmer, Helmut (2007): Language Across the Curriculum – A Way towards Plurilingualism. In: Martyniuk, Waldemar (Hrsg): Towards a Common European Framework of Reference for Languages of School Education. Krakow: Universitas, 177-192.
- Wokusch, Susanne / Lys, Irène (2007): Überlegungen zu einer integrativen Fremdsprachendidaktik. In: Beiträge zur Lehrerbildung 25/2, 168-179.
- Wolff, Dieter (1998): Lernstrategien: Ein Weg zu mehr Lernerautonomie. Online: <http://paedpsych.jk.uni-linz.ac.at:4711/LEHRTEXTE/Wolff98.html> (01.2002).
- Wenden, Anita / Rubin, Joan (Hrsg.) (1987): Learner Strategies in Language Learning. Englewood Cliffs: Prentice-Hall.

5 Propositions des experts pour des standards de base

Le chapitre 3.6 du présent rapport compare des compétences en lecture et en expression écrite, mesurées dans le cadre des enquêtes HarmoS, avec les propositions de standards de base avancées par les experts. Il apparaît, dans l'ensemble – avec quelques légères différences – que seulement près de 50% des élèves actuellement en sixième et en neuvième année atteindraient les standards de formation suisses, définis comme des standards de base, c'est-à-dire des standards minimaux. Or les standards de base devraient, une fois qu'ils ont été adoptés, être en fait atteints par tous les élèves des niveaux concernés, pour autant que ceux-ci ne connaissent pas de graves difficultés d'apprentissage ou des circonstances exceptionnellement défavorables, qui les empêcheraient d'atteindre les standards fixés.

Le consortium Langues étrangères s'est trouvé dans une situation difficile pour donner un avis d'experts, puisqu'il a dû prédire, sur la base des compétences mesurées auprès d'élèves dans le courant de l'année 2007, les compétences (minimales) des élèves qui, au moment des prochaines mesures, dans quelques années, auront suivi des cours dans des conditions tout à fait autres, dont on pense qu'elles permettront d'obtenir de meilleurs résultats.

En vue d'étayer plus largement sa proposition de standards de base, le consortium Langues étrangères a organisé, les 23 et 24 novembre 2007, une rencontre d'experts (panel) sur la définition des standards de base. Y ont participé les membres et le conseil scientifique du consortium, ainsi que d'autres experts actifs dans le domaine de la didactique des langues étrangères – en tout, plus de 20 personnes.

De manière générale, on a travaillé à partir de descriptions de niveaux, d'exemples d'exercices et de performances d'élèves. A cela se sont ajoutées des informations sur les performances des élèves, collectées à l'occasion des tests HarmoS – particulièrement en lien avec la compréhension de l'écrit. Concrètement, le processus de définition s'est déroulé selon le schéma suivant pour toutes les aptitudes:

- 1) Les participants au panel ont examiné le matériel mentionné, puis ont dû décider provisoirement, à partir des descriptions de compétences (base: *Cadre de référence* et *IEF / lingualevel*), quel était le niveau minimal qui devrait être atteint par pratiquement tous les élèves des futures classes de sixième et de neuvième années.
- 2) Ensuite, on leur a présenté des exercices et des performances qui, de l'avis du consortium, illustraient ces niveaux. La première question à laquelle les participants devaient répondre était celle de savoir si les exercices / performances présentés pouvaient illustrer le niveau en question.
- 3) Dès que les participants s'entendaient sur les performances qui convenaient, ils avaient l'occasion de revenir sur leur décision concernant les niveaux et d'ouvrir le débat sur le niveau qui leur semblait être «le bon».

4) La décision quant au niveau à fixer pour chaque standard de base a finalement été prise à l'aide d'un vote (à main levée).

Cette façon de procéder n'a pu être suivie dans toutes ses étapes pour chacune des aptitudes (pas pour la compréhension de l'oral, par ex.); mais toutes les suggestions ont été débattues sur la base au moins des descriptions de compétences, puis des décisions prises en conséquence.

Le consensus au sein du groupe était assez élevé. Il n'y a eu que quelques rares avis divergents, et ce tant vers le bas que vers le haut. Au fil de la rencontre, il est apparu assez rapidement que, pour la grande majorité des experts, il ne s'agit pas de maintenir la pratique actuelle dans l'enseignement des langues étrangères, mais bien de saisir la chance qu'offrent les standards de formation pour faire un pas décisif en avant et adopter, dans les langues étrangères, un nouveau type d'enseignement, qui soit plus près de la pratique et plus à même de mener les élèves vers le but à atteindre.

Lors de la rencontre, les experts ont régulièrement été confrontés à la difficulté qui consistait à penser à des standards de base uniquement et systématiquement. Dans les remarques des participants, il apparaît qu'ils étaient souvent tentés de penser à «l'élève moyen» et à ce que ce dernier devrait être capable de faire au terme d'un niveau donné.

Afin de faciliter la compréhension du lecteur, nous reproduisons ci-dessous l'aperçu des décisions prises par le panel d'experts au sujet des niveaux à atteindre. Il s'agit d'une vue d'ensemble de la proposition de standards complète avancée par les experts:

Standard niveau 8^e année / 1^{re} langue étrangère

Niveaux de référence		A1.1	A1.2	A2.1	A2.2	B1.1	B1.2	B2
Niveaux HarmoS		I ₈	II ₈	III ₈	IV ₈	V ₈		
Domaines de compétences	Compréhension de l'oral	→						
	Compréhension de l'écrit	→						
	Prendre part à une conversation	→						
	S'exprimer oralement en continu	→						
	Ecrire	→						
	Médiation linguistique	→		→				

Standard niveau 8^e année / 2^e langue étrangère

Niveaux de référence		A1.1	A1.2	A2.1	A2.2	B1.1	B1.2	B2
Niveaux HarmoS		I ₈	II ₈	III ₈	IV ₈	V ₈		
Domaines de compétences	Compréhension de l'oral	→						
	Compréhension de l'écrit	→						
	Prendre part à une conversation	→						
	S'exprimer oralement en continu	→						
	Ecrire	→						
	Médiation	→						

	linguistique							
--	--------------	--	--	--	--	--	--	--

Standard niveau 11^e année: 1^{re} et 2^e langues étrangères

Niveaux de référence		A1.1	A1.2	A2.1	A2.2	B1.1	B1.2	B2
Niveaux HarmoS		I ₁₁			II ₁₁	III ₁₁	IV ₁₁	V ₁₁
Domaines de compétences	Compréhension de l'oral	→						
	Compréhension de l'écrit	→						
	Prendre part à une conversation	→						
	S'exprimer oralement en continu	→						
	Ecrire	→						
	Médiation linguistique	→			→			

III. 23: Aperçu des propositions des experts pour des standards de base dans les disciplines des langues étrangères

Le consortium approuve la proposition avancée par les experts, mais il la voit également comme une contribution au débat; il est donc prêt à en débattre ouvertement, sans prévention.

6 Références bibliographiques

La liste des références bibliographiques comprend les titres déjà mentionnés avec les projets présentés et des titres qui se sont ajoutés aux références au fil du projet. La littérature spécialisée sur les compétences transversales – séparée en compétences interculturelles et méthodologiques – a été intégrée à la section 4.2.4. Quant aux titres qui sont aussi importants pour des sujets en dehors de ce domaine thématique restreint, il ont pour certains été repris ci-après, dans la liste des références bibliographiques générales.

- Adamzik, Kirsten / Antos, Gerd / Jakobs, Eva-Maria (Hg.) (1997): Domänen- und kulturspezifisches Schreiben. Bern: Lang (= Textproduktion und Medium 3).
- Alderson, Charles J. / Banerjee, Jayanti (2001/2002): State-of-the-Art Review: Language testing and assessment. In: Language Teaching 2001 (Part 1) / 2002,4 (Part 2).
- Alderson, Charles J. / Clapham, Caroline / Wall, Dianne (1995): Language Test Construction and Evaluation. Cambridge: Cambridge University Press (= Language Teaching Library).
- Alderson, J. Charles (2000): Assessing Reading. Cambridge: CUP (= Language Assessment).
- Alderson, J. Charles (2004). The Dutch CEF Construct Group: advice on how to develop specifications for tests at each level of the CEF. Online: <http://www.ealta.eu.org/conference/2004/ppt/alderson14may2.ppt> (20.07.08).
- Alderson, J.C. / Figueras, N. / Kuijper, H. / Nold, G. / Takala, S. / Tardieu, C. (2004): The Development of Specifications for Item Development and Classification within The Common European Framework of Reference for Languages: Learning, Teaching, Assessment: Reading and Listening. Online: <http://www.lancs.ac.uk/fss/rojects/grid/> (20.07.08).
- ALTE/Association of Language Testers in Europe (o.J. a): The CEFR Grid for Writing Tasks v. 3.1 (analysis) Online: http://www.coe.int/T/DG4/Portfolio/documents/CEFRWritingGridv3_1_analysis.doc (25.01.2007).
- ALTE/Association of Language Testers in Europe (o.J. b): The CEFR Grid for Writing Tasks v. 3.1 (presentation) Online: http://www.coe.int/T/DG4/Portfolio/documents/CEFRWritingGridv3_1_presentation.doc (25.01.2007).

- Altmann, Gerry T. M. (1990): Cognitive models of speech processing. Cambridge, Mass: MIT Press.
- Bachman, Lyle F. (1990): Fundamental considerations in language testing. Oxford: Oxford University Press.
- Bachman, Lyle F. (2000): Modern Language testing at the turn of the century: assuring that what we count counts. In: Language Testing 17 / 1, 1-42.
- Bachman, Lyle F. (2002): Some reflections on task-based language performance assessment. In: Language Testing 19, (3), 453-476.
- Bachman, Lyle F. (2004): Statistical analyses for language assessment. Cambridge: CUP.
- Bachman, Lyle F. / Cohen, Andrew D. (Hg.) (1998): Interfaces Between Second Language Acquisition and Language Testing Research. Cambridge: University Press (= Applied Linguistics).
- Bachman, Lyle F. / Palmer, Adrian S. (1996): Language Testing in Practice. Designing and Developing Useful. Language Tests. Oxford: University Press.
- Baden-Württemberg, Ministerium für Kultus, Jugend und Sport Baden-Württemberg (Hg.) (2004). Bildungspläne. Online: <http://www.bildungsstandards-bw.de/>
- Baer, Matthias / Jurt, Ueli / Fuchs, Michael, Reber-Wyss, Monika, und Nussbaum, Thomas (1997): Wie unterschiedlich gute und verschieden alte Textverfasser/innen Schreibaufgaben bewältigen und Möglichkeiten, Schüler/innen beim Aufbau der Schreibkompetenz zu unterstützen. St. Gallen: UVK Fachverlag für Wissenschaft und Studium.
- Banerjee, J. (2004): Qualitative analysis methods. In: Council of Europe (Hg.): Reference Supplement to the Preliminary Pilot version of the Manual for Relating Language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment, Section D. Strasbourg: Council of Europe (DGIV/EDU/LANG (2004) 13). Online: www.coe.int/portfolio > Documentation (26.01.2007).
- Bausch, Karl-Richard / Burwitz-Melzer, Eva / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.) (2006): Aufgabenorientierung als Aufgabe. Tübingen: Narr.
- Bausch, Karl-Richard / Burwitz-Melzer, Eva / Königs, Frank G. (Hrsg., 2005): Bildungsstandards für den Fremdsprachenunterricht auf dem Prüfstand. Tübingen: Narr.
- Bausch, Karl-Richard / Christ, Herbert / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.) (2002): Der Gemeinsame europäische Referenzrahmen in der Diskussion. Tübingen: Narr.
- Beacco, Jean-Claude (2004): Une proposition de référentiel pour les compétences culturelles dans les enseignements de langues. In : Beacco, Jean-Claude/ Bouquet, Simon/ Porquier, Rémy (2004): Niveau B2 pour le Français. Textes et Références. Paris: Didier, 251-286.
- Beacco, Jean-Claude / Byram, Michael (2003): Guide pour l'élaboration des politiques linguistiques éducatives en Europe: de la diversité linguistique à l'éducation plurilingue. Version intégrale. Projet1. Strasbourg: Conseil de l'Europe.. http://www.coe.int/T/E/Cultural_Co-operation/education/Languages/Language_Policy
- Beck, Bärbel / Klieme, Eckhard (2003): DESI: Eine Längsschnittstudie zur Untersuchung des Sprachunterrichts in deutschen Schulen. In: Finkbeiner, C. / Fehling, S. (Hg.): Evaluation im Brennpunkt - Fremdsprachen lernen und lehren 17 / 3, 380-395.
- Beck, Bärbel / Klieme, Eckhard (Hrsg., 2007): Sprachliche Kompetenzen. Konzepte und Messung. DESI-Ergebnisse Band 1. Weinheim und Basel: Beltz.
- Bereiter, Carl / Scardamalia, Marlene (1987): The Psychology of Written Composition. Hillsdale: Erlbaum.
- Blue, George M. / Milton, James / Saville, Jane (Hg.) (2000): Assessing English for Academic Purposes. Bern: Lang.
- Bolton, Sibylle (2003): Qualitätsmanagement bei den Prüfungen des Association of Language Testers in Europe (ATE). In: AKS-Rundbrief, Fremdsprachen und Hochschule. Thema: Qualitätssicherung im Sprachtesten 69, 9-22.
- Bond, Trevor G. / Fox, Christine M. (2001): Applying the Rasch Model. Fundamental Measurement in the Human Sciences. Mahwah, N.J.; London: Erlbaum.
- Börner, Wolfgang / Vogel, Klaus (Hg.) (1992): Schreiben in der Fremdsprache. Prozess und Text, Lehren und Lernen. Bochum: AKS (= FLF Fremdsprachen in Lehre und Forschung 10).
- Bregy, Anne-Lore / Revaz, Nadia (2001): Evaluation des enseignements / apprentissages bilingues en classes de 4ème primaire de Sierre. Année scolaire 1999 / 2000. Rapport intermédiaire. Neuchâtel: IRDP.
- Brindley, Geoff (2001): Outcomes-based assessment in practice: some examples and emerging insights. In: Language Testing 18, 4.
- Brohy, Claudine (1999): L'enseignement bilingue / l'enseignement précoce des langues secondes en Suisse romande. Neuchâtel: IRDP (= IRDP Regards).
- Brohy, Claudine / Pannatier, Monique (2000): L' évaluation dans l' enseignement immersif - la quadrature du cercle? In: Babylonia. Themenheft: Lernerfahrungen reflektieren - Lernerfolge beurteilen. Réfléchir sur les expériences d' apprentissage - en évaluer les résultats 1, 33-35.
- Brown, Gillian / Malmkjaer, Kirsten / Pollitt, Alastair / Williams, John (eds.) (1994): Language and understanding. Oxford: University Press (= Applied Linguistics).
- Brown, Gillian / Malmkjaer, Kirsten / Williams, John (Hg.) (1996): Performance and Competence in Second Language Acquisition. Cambridge: University Press.

- Brown, James Dean / Hudson, Thom (2002): *Criterion-referenced Language Testing*. Cambridge: University Press.
- Buck, Gary (2001): *Assessing Listening*. Cambridge: University Press.
- Burwitz-Melzer, Eva (2006): *Interkulturelles und sprachliches Lernen mit fremdsprachlichen literarischen Texten: Zwei zentrale Elemente eines neuen Lesekompetenzmodells*. In: *FLUL* 35, 104-120.
- Bygate, Martin / Skehan, Peter / Swain, Merrill (2001): *Researching Pedagogic Task*. Second Language Learning, Teaching and Testing. Harlow: Pearson Education.
- Byram, Michael (1997): *Teaching and Assessing Intercultural Communicative Competence*. Clevedon: Multilingual Matters.
- Byram, Michael / Neuner, Gerhard / Zarate, Geneviève (1997): *La compétence socioculturelle dans l'apprentissage et l'enseignement des langues. Vers un Cadre Européen de référence pour l'apprentissage et l'enseignement des langues vivantes: études préparatoires*. Strasbourg: Editions du Conseil de l'Europe (= Education).
- Byram, Michael / Nichols, Adam / Stevens, David (Hg.) (2001): *Developing intercultural competence in practice*. Clevedon: Multilingual Matters.
- Byram, Michael / Tost Planet, Manuel (Hg.) (2000): *Identité sociale et dimension européenne: la compétence interculturelle par l'apprentissage des langues vivantes*. Graz / Strasbourg: European Centre for Modern Languages / Conseil de l'Europe.
- Byram, Michael / Tost Planet, Manuel (Hg.) (2000): *Sociale Identity and European Dimension: Intercultural Competence Through Foreign Language Learning*. Graz / Strasbourg: European Centre for Modern Languages / Conseil de l'Europe.
- Byram, Michael / Zarate, Geneviève (Hg.) (1997): *The sociocultural and intercultural dimension of language learning and teaching*. Strasbourg: Council of Europe Publishing (= Education).
- Canale, M. / Swain, M. (1980): *Theoretical bases of communicative approaches to second language teaching and testing*. In: *Applied Linguistics* 1 (1), 1-47.
- Candlin, Christopher N. / Hyland, Ken (eds.) (1999): *Writing: texts, processes and practices*. London / New York: Longman (= Applied Linguistics and Language Study).
- Caspari, Daniela (2005): *Von Bildungsstandards zu Lehrplänen*. In: Bausch, K.-R. / Burwitz-Melzer, E. / Königs, F.G. / Krumm, H.-J. (Hrsg.): *Bildungsstandards für den Fremdsprachenunterricht auf dem Prüfstand*. Tübingen, 67-77.
- Chapelle, C. (1998): *Construct definition and validity inquiry in SLA research*. In: Bachman, L. / Cohen, A. (Hg.): *Interfaces between second language acquisition and language testing research*. Cambridge: CUP, 32-70.
- Chapelle, C. / Grabe, W. / Berns, M. (1997): *Communicative language proficiency: definitions and implications for TOEFL 2000*. Monograph Series. Princeton, N.J.: ETS.
- Chardenet, Patrick (1999): *De l'activité évaluative à l'acte d'évaluation. Approche théorique et pratique communicationnelle*. Paris: L'Harmattan (= Savoir et formation).
- Charge, Nick / Williams, Andy (2003): *IELTS - Quality Assurance in English Language Testing*. In: *AKS-Rundbrief, Fremdsprachen und Hochschule*. Thema: Qualitätssicherung im Sprachtesten 69, 23-27.
- Charpié, Nathalie (2002): *Expérimentation de l'enseignement de l'allemand par immersion dans le canton du Jura*. Neuenburg: IRDP (= IRDP document de travail).
- Christ, Ingeborg (2003): *Auf dem Wege zu einer neuen Evaluationskultur im Fremdsprachenunterricht*. In: *Neusprachliche Mitteilungen aus Wissenschaft und Praxis* 3, 157-169.
- CIIP (2003). *Déclaration de la CIIP relative à la politique de l'enseignement des langues en Suisse romande*. Online: www.ciip.ch/ciip/pdf/cp030403_3%20.pdf
- Clapham, Caroline / Corson, David (eds.) (1997): *Encyclopedia of Language and Education, Volume 7: Language Testing and Assessment*. Dordrecht: Kluwer Academic Publishers.
- Conseil de l'Europe (2001). *Cadre européen commun de référence pour les langues: apprendre, enseigner, évaluer*. Strasbourg: Conseil de l'Europe / Didier (= Apprentissage des langues et citoyenneté européenne).
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: CUP. Online: www.coe.int/portfolio >Documentation
- Council of Europe (2002a). *Cadre européen commun de référence pour les langues: apprendre, enseigner, évaluer (CECR). Guide pour les utilisateurs*. Strasbourg: Conseil de l'Europe (= Division des Politiques Linguistiques).
- Council of Europe (2002b). *Common European Framework of Reference for Languages: Learning, teaching, assessment. Case studies* (Hg. Ch. Alderson). Strasbourg: Council of Europe (= Language Policies).
- Council of Europe (2003). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEF). Manual, Preliminary Pilot Version*. Strasbourg: Language Policies. Online: http://www.coe.int/T/DG4/Linguistic/Manuel1_EN.asp. (20.07.2008).
- Council of Europe (2009). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A Manual*. Strasbourg: Language Policy Division.

- Council of Europe (2004). Relating language examinations to the Common European framework of reference for languages (CEF). Reference supplement to the preliminary pilot version of the manual. Strasbourg: Language Policies. Online: http://www.coe.int/T/DG4/Linguistic/Manuel1_EN.asp. (20.07.2008).
- Cushing Weigle, Sara (2002): *Assessing Writing*. Cambridge: University Press (= Cambridge Language Assessment Series).
- Davidson, Fred / Lynch, Brian K. (2002): *Testcraft: a teacher's guide to writing and using language test specifications*. New Haven: Yale University Press.
- Davies, A. (1989): Communicative competence as language use. In: *Applied Linguistics*, 10 (2), 157-170.
- Dayez, Yves (2001): *Guide des sujets du DELF et du DALF. Connaître, concevoir, évaluer*. Paris: Didier (= Les cahiers du CIEP).
- de Pietro, Jean-François / Lüdi, George (Hg.) (1991): *Modèles d'évaluation du français langue seconde dans l'éducation publique. Dossier établi par le département de linguistique française du Séminaire des langues romanes de l'Université Bâle*. Basel.
- Demierre-Wagner, Andrea / Schwob, Irène / Ducrey, François (2004): L'expérience pilote de l'enseignement bilingue à l'école primaire en Valais romand. In: *bulletin vals-asla* 79.
- Depover, Christian / Noël, Bernadette (éds) (1999): *L' évaluation des compétences et des processus cognitifs. Modèles, pratiques et contextes*. Bruxelles: De Boeck (= Pédagogies en développement).
- Douglas, Dan (2000): *Assessing Languages for Specific Purposes*. Cambridge: CUP (= Language Assessment).
- EDK, Schweizerische Konferenz der kantonalen Erziehungsdirektoren (2004). *Stratégie de la CDIP et programme de travail pour la coordination de l'enseignement des langues à l'école obligatoire en Suisse*. En ligne: <http://edudoc.ch/getfile.py?docid=1685&name=7-2f&format=pdf&version=1> (20.07.08).
- EDK, Schweizerische Konferenz der kantonalen Erziehungsdirektoren, Expertengruppe (1998). *Quelles langues apprendre en Suisse pendant la scolarité obligatoire ?* En ligne : <http://sprachenkonzept.franz.unibas.ch/concept.html>
- EDK, Schweizerische Konferenz der kantonalen Erziehungsdirektoren (2004). *HARMOS. Zielsetzungen und Konzeption*. Bern: EDK Generalsekretariat, Koordinationsbereich „Obligatorische Schule“. Online: <http://www.edk.ch/dyn/11968.php> (20.07.08).
- EDK, Schweizerische Konferenz der kantonalen Erziehungsdirektoren (2001). *Recommandations relatives à la coordination de l'enseignement des langues au niveau de la scolarité obligatoire*. Projet du 1.6.2001
- Ehlers, Swantje (2006): Entwicklung von Lesekompetenz in der Fremdsprache. In: *Babylonia* 3-4, 31-38.
- Ellis, Rod (2000): Task-based research and language pedagogy. In: *Language Teaching Research* 4,3 (2000), 193–220.
- Ellis, Rod (2003): *Task-based Language and Teaching*. Oxford: OUP.
- Ellis, Rod (Hrsg.) (2005): *Planning and Task Performance in a Second Language*. Amsterdam u. a.: John Benjamins.
- Ellis, Rod / Barkhuizen, Gary (2005): *Analysing Learner Language*. Oxford: OUP.
- Engel, Ulrich (1988): *Deutsche Grammatik* (2. Aufl.). Heidelberg: Julius Groos.
- Eriksson, Brigit / Le Pape Racine, Christine / Reutener, Hans / Serra, Cecilia / Stern, Otto (1998): *Français - Allemand: Apprendre en deux langues à l' école secondaire. Rapport de valorisation*. Aarau: CSRE (= L' efficacité de nos systèmes de formation).
- Eriksson, Brigit / Le Pape Racine, Christine / Reutener, Hans / Serra, Cecilia / Stern, Otto (1999): *Französisch - Deutsch: Zweisprachiges Lernen auf der Sekundarstufe I*. Chur / Zürich: Rüegger (= Wirksamkeit unserer Bildungssysteme. NFP 33).
- Europarat (2001). *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Berlin u. a.: Langenscheidt. Online: <http://www.goethe.de/referenzrahmen> (30.05.08)
- Finkbeiner, C. / Fehling, S. (Hg.) (2003): *Evaluation im Brennpunkt - Fremdsprachen lernen und lehren*. Landau: Empirische Pädagogik (= Empirische Pädagogik 17 / 3).
- Fuchs, Gabriela / Werlen, Iwar (1995). *Zwischenbericht zum Projekt "Zweisprachige Schule Brig-Glis" an den Schulpräsidenten zuhänden des Gemeinderates des Stadtgemeinde Brig-Glis*. Brig-Glis/ Bern: Universitäres Forschungszentrum für Mehrsprachigkeit.
- Fuchs, Gabriela / Werlen, Iwar (1996). *Zweisprachige Schule Brig-Glis. Zwischenbericht zum Projekt "Zweisprachige Schule Brig-Glis" an den Schulpräsidenten zuhänden des Gemeinderates der Stadtgemeinde Brig-Glis*. Brig-Glis, Universitäres Forschungszentrum für Mehrsprachigkeit.
- Fuchs, Gabriela / Werlen, Iwar (1997): *Zweisprachige Schule Brig-Glis. Zwischenbericht nach zwei Jahren teilimmersivem Unterricht. Schuljahr 1995/96 und 1996/97. Bericht an den Schulpräsidenten zuhänden des Gemeinderates der Stadtgemeinde Brig-Glis*. Brig-Glis: Universitäres Forschungszentrum für Mehrsprachigkeit.
- Fuchs, Gabriela / Werlen, Iwar (1998): *Zweisprachige Schule Brig-Glis. Schuljahr 1997/98 Kurzfassung*. Bern, Brig-Glis: Universitäres Forschungszentrum für Mehrsprachigkeit.
- Fulcher, Glenn (2003): *Testing Second Language Speaking*. London: Longman / Pearson Education (= Applied Linguistics and Language Study).

- Garman, Mike (1990): Psycholinguistics. Cambridge: Cambridge University Press.
- Gassner, Otmar / Horak, Angela / Mewald, Claudia / Moser, Wolfgang / Schober, Michael / Stefan, Ferdinand / Valsky, Claudia (2005): Bildungsstandards in Österreich. Fremdsprachen. Englisch 8. Schulstufe. Version September'05. Wien: Bundesministerium für Bildung, Wissenschaft und Kultur.
- Gernsbacher, Morton Ann (Hg.) (1994): Handbook of psycholinguistics. San Diego: Academic Press.
- Gernsbacher, Morton Ann / Givón, Talmy (Hg.) (1995): Coherence in spontaneous text. Amsterdam: John Benjamins (=Typological studies in language; 31).
- Gick, Cornelia / Langner, Michael / Lenz, Peter / Schneider, Günther (1993): Untersuchungen zur Evaluation der mündlichen und schriftlichen Kommunikationsfähigkeit. In: Dossier, EDK: Analyse von Evaluationsmodellen für Deutsch und Französisch L2 am Ende der obligatorischen Schulzeit 23/2, 238-322.
- Glaboniat, Manuela (1998): Kommunikatives Testen im Bereich Deutsch als Fremdsprache. Eine Untersuchung am Beispiel des Österreichischen Sprachdiploms Deutsch. Innsbruck / Wien: Studien (= Theorie und Praxis - Österreichische Beiträge zu Deutsch als Fremdsprache, Serie B 2).
- Glaboniat, Manuela / Müller, Martin / Rusch, Paul / Schmitz, Helen / Wertenschlag, Lukas (2005): Profile deutsch. Gemeinsamer europäischer Referenzrahmen. Lernzielbestimmungen - Kannbeschreibungen - Kommunikative Mittel - Niveau A1 - A2 - B1 - B2 - C1 - C2 mit CD-ROM. München: Langenscheidt.
- Gajo, Laurent / Matthey, Marinette / Py, Bernard (1999): L'école, les copains, la famille... Les apprentissages linguistiques des enfants de migrants dans une région de Suisse romande, Berne/Aarau, PNR33/CSRE, 1997 (=rapport de valorisation)
- Grabe, William / Kaplan, Robert (1996): Theory and practice of writing: an applied linguistic perspective. London: Longman (= Applied linguistics and language study).
- Gregori, Gian Peder / Todisco, Vincenzo (2003). Evaluation des Zweitsprachunterrichts (ZSU) Italienisch in den Primarschulen des deutschsprachigen Teils des Kantons Graubünden. Schlussbericht. Chur, Pädagogische Fachhochschule Graubünden.
- Grotjahn, Rüdiger (1995): Der C-Test: State of Art. In: Zeitschrift für Fremdsprachenforschung (ZFF) 6/2, 37-60.
- Grotjahn, Rüdiger (1995): Zweitsprachliches Leseverstehen: Grundlagen und Probleme der Evaluation. In: Die Neueren Sprachen 5, 533-555.
- Grotjahn, Rüdiger (1997): Der C-Test: Neuere Entwicklungen. In: Gardenghi, Monica / O'Connell, Mary: Prüfen, Testen, Bewerten im modernen Fremdsprachenunterricht 6, 117-128.
- Grotjahn, Rüdiger (Hg.) (2002): Der C-Test. Theoretische Grundlagen und praktische Anwendungen. Band 4. Bochum: AKS (= FLF Fremdsprachen in Lehre und Forschung 32).
- Grotjahn, Rüdiger (2000): Determinanten der Schwierigkeit von Leseverstehensaufgaben: Theoretische Grundlagen und Konsequenzen für die Entwicklung des TESTDAF. In: Bolton, Sybille, Hrsg., TESTDAF: Grundlagen für die Entwicklung eines neuen Sprachtests. Beiträge aus einem Expertenseminar. München: Goethe-Institut.
- Habel, Christopher et al. (Hg.) (1996): Perspektiven der kognitiven Linguistik: Modelle und Methoden. Opladen: Westdeutscher Verlag (=Psycholinguistische Studien).
- Hamp-Lyons, Liz (Hg.) (1991): Assessing second language writing in academic contexts. Norwood: Ablex Publishing Corporation.
- Harsch, Claudia (2006): Der Gemeinsame europäische Referenzrahmen: Leistung und Grenzen. Die Bedeutung des Referenzrahmens im Kontext der Beurteilung von Sprachvermögen am Beispiel des semikreativen Schreibens im DESI-Projekt. Inaugural-Dissertation. Online: http://www.opus-bayern.de/uni-augsburg/frontdoor.php?source_opus=368 (20.07.08).
- Hayes, John R. / Flower, Linda S. (1980): Identifying the Organization of Writing Processes. In: Gregg, Lee W./Steinberg, Edwin R. (Hrsg.): Cognitive processes in writing. Hillsdale: Erlbaum, 3-30.
- Henrici, Gert / Zöfgen, Ekkehard (Hg.) (2001): Themenheft: Leistungsmessung und Leistungsevaluation. In: Fremdsprachen lehren und lernen (FLuL) 30.
- Herrmann, Theo / Grabowski, Joachim (1995): Sprechen: Psychologie der Sprachproduktion. Heidelberg: Spektrum.
- Hyland, Ken (2002): Teaching and Researching Writing, Applied Linguistics in Action. Harlow: Pearson Education (= Applied Linguistics in Action).
- Hyland, Ken (2003): Second Language Writing. Cambridge: ELT (= Language Education).
- Ioannou-Georgiou, Sophie / Pavlou, Pavlos (2003): Assessing Young Learners. Oxford: University Press (= Ressource books for Teachers).
- IRDP (1998): Le point sur la recherche. Enseignement bilingue. Bulletin d'information janvier 1998. In: Mehrsprachigkeit.
- Jones, Neil (2002). Relating the ALTE Framework to the Common European Framework of reference. In Council of Europe. Common European Framework of Reference for Languages: Learning, teaching, assessment. Case studies (Hg. Ch. Alderson). Strasbourg: Council of Europe, 167 – 183. Online: http://www.coe.int/T/DG4/Portfolio/documents/case_studies_CEF.doc (20.07.08).

- Kaftandjieva, Felianka / Takala, Sauli (2002). Council of Europe Scales of Language Proficiency: A validation study. In Council of Europe. Common European Framework of Reference for Languages: Learning, teaching, assessment. Case studies. (Hg. Ch. Alderson). Strasbourg: Council of Europe, 106 – 129. Online: http://www.coe.int/T/DG4/Portfolio/documents/case_studies_CEF.doc (20.07.08).
- Kästner, Uwe (1997): Freies Schreiben in der Fremdsprache - Prozesse und ihre Didaktik. Untersuchung zur Textproduktion im fremdsprachlichen Deutschunterricht auf dem Hintergrund neuropsychologischer Theorien. Bochum: AKS (= FLF Fremdsprachen in Lehre und Forschung 19).
- Kintsch, Walter (1998): Comprehension. A paradigm of cognition. Cambridge: Cambridge University Press.
- Klieme, Eckhard et. al.. (2003): Le développement de standards nationaux de formation. Une expertise. Bonn: Ministère fédéral de l'Education et de la Recherche (BMBF). En ligne : http://www.bmbf.de/pub/le_developpement_de_standards_nationaux_d_formation.pdf
- KMK, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hg.) (2004). Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Mittleren Schulabschluss. Neuwied: Luchterhand. Online: <http://www.kmk.org/doc/beschl/aschulw.htm> (25.06.08)
- Kniffka, Gabriele / Üstünsöz-Beurer, Dörthe (2001): TestDaF: Mündlicher Ausdruck. Zur Entwicklung eines kassettengesteuerten Testformats. In: Henrici, Gert / Zöfgen, Ekkehard (Hg.): Fremdsprachen lehren und lernen (FLuL) 30, 127-149.
- Knobloch, Clemens (1994): Sprache und Sprechfähigkeit: Sprachpsychologische Konzepte. Tübingen: Niemeyer.
- Kobayashi, Miyoko (2002): Method effects on reading comprehension test performance: text organization and response format. In: *Language Testing*, 19, 193-220.
- Krause, Wolf-Dieter / Sändig, Uta (2002): Testen und Bewerten kommunikativer Leistungen im Unterricht Deutsch als Fremdsprache. Linguistische Grundlagen und didaktische Angebote. Frankfurt/M. u. a.: Lang (= Sprache - System und Tätigkeit 43).
- Krumm, Hans-Jürgen (Hg.) (1999): Erfahrungen beim Schreiben in der Fremdsprache Deutsch. Untersuchungen zum Schreibprozess und zur Schreibförderung im Unterricht mit Studierenden. Innsbruck / Wien: Studien (= Theorie und Praxis - Österreichische Beiträge zu Deutsch als Fremdsprache, Serie B 4).
- Kunnan, Antony John (ed.) (2000): Fairness and Validation in Language Assessment: Selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida. Cambridge: University Press (= Studies in Language Testing 9).
- Kurtz, Jürgen (2001): Improvisierendes Sprechen im Fremdsprachenunterricht. Eine Untersuchung zur Entwicklung spontansprachlicher Handlungskompetenz in der Zielsprache, Tübingen: Gunter Narr Verlag (gnv). Labudde, Peter (Hrsg.) (2007): Bildungsstandards am Gymnasium. Korsett oder Katalysator? Bern: hep / Pädagogische Hochschule.
- Labudde, Peter (Hrsg.) (2007): Bildungsstandards am Gymnasium. Korsett oder Katalysator? Bern: hep / Pädagogische Hochschule.
- Language Policy Division / Division des Politiques Linguistiques (eds.) (2004): Relating language examinations to the CEF Reference reading items and writing tasks. Preliminary pilot version of the Manual and related documents. Relier les examens de langues au CECR. Items et tâches de référence pour des tests de lecture et d'écriture. Avant-projet de Manuel et documents corrélatifs. Strasbourg: Conseil de l'Europe. http://www.coe.int/T/DG4/Portfolio?L=E&M=/main_pages/illustrationse.html (20.07.08)
- Language Testing (2000). Special Issue on Assessing Young Language Learners. Vol. 17, iss. 2.
- Lenz, Peter (2000): Erfahrungen mit dem Europäischen Sprachenportfolio in der Schweiz. In: *Babylonia*. Themenheft: Lernerfahrungen reflektieren - Lernerfolge beurteilen. Réfléchir sur les expériences d'apprentissage - en évaluer les résultats 1, 23-28.
- Lenz, Peter (2002): Mehrsprachigkeit fördern und sichtbar machen - das Europäische Sprachenportfolio. In: Schneider, Günther / Clalüna, Monika (Hg.): Bulletin suisse de linguistique appliquée. VALS/ASLA. Sonderheft: Mehrsprachigkeit und Deutschunterricht: Thesen, Beiträge und Berichte aus der Sektionsarbeit an der XII. IDT - Internationalen Tagung der Deutschlehrerinnen und Deutschlehrer 2001 in Luzern Sonderheft, 151-159.
- Lenz, Peter (2004a): The European Language Portfolio - A Tool for Comprehensive Assessment and the Promotion of Plurilingualism. In: Bräuer, Gerd / Sanders, Karen (ed): *New Visions in Foreign and Second Language Education*, 123-152.
- Lenz, Peter (2004b): The European Language Portfolio. In: Morrow, Keith (Hg.): *Insights from the Common European Framework*. Oxford: Oxford University Press, S. 22-31.
- Lenz, Peter (2006): Überlegungen zur Sprachkompetenzbeschreibung und Testvalidierung im Projekt HarMoS/Fremdsprachen. In: Mondada, Lorenza / Pekarek Doehler, Simona (Hrsg.): *Bulletin suisse de linguistique appliquée*. VALS / ASLA. La notion de compétence: études critiques 84, 191-227.
- Lenz, Peter / Schneider, Günther (2002): Developing the Swiss model of the European Language Portfolio. In: Council of Europe / Alderson, J.C. (Hg.): *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Case Studies*. Strasbourg: Council of Europe Publishing 2002, 68-86.
- Lenz, Peter / Schneider, Günther (2004a). Introduction to the bank of descriptors for self-assessment in European Language Portfolios. Strasbourg: Council of Europe, Language Policy Division. Online: www.coe.int/portfolio (01.02.08)

- Lenz, Peter / Schneider, Günther (2004b). A bank of descriptors for self-assessment in European Language Portfolios. Strasbourg: Council of Europe, Language Policy Division. Online: www.coe.int/portfolio (01.02.08)
- Lenz, Peter/ Studer, Thomas (2004): Sprachkompetenzen von Jugendlichen einschätzbar machen. In: *Babylonia*. 2, 21-25.
- Lenz, Peter; Studer, Thomas (2007): *lingualevel*. Instrumente zur Evaluation von Fremdsprachenkompetenzen. Hrsg. v. Bildungsdirektoren-Konferenz Zentralschweiz (BKZ) / Nordwestschweizerische Erziehungsdirektoren-Konferenz (NW-EDK) / Erziehungsdirektoren-Konferenz der Ostschweizer Kantone und des Fürstentums Liechtenstein (EDK-Ost). Bern: Schulverlag bmv.
- Levelt, Willem (1989): *Speaking. From Intention to Articulation*. Cambridge MA: Bradford Book.
- Linacre, J. M. (2007). A user's guide to Facets: Rasch-measurement computer program. Chicago. Online: www.winsteps.com/facets.htm (01.02.08)
- Linacre, J. M. (2007). A user's guide to WINSTEPS-MINISTEP: Rasch-model computer programs. Chicago. Online: www.winsteps.com/winsteps.htm (01.02.08).
- Linacre, John M. (1992). *FACETS: A Computer Program for the Analysis of Multi-Faceted Data*. Chicago: MESA Press.
- Little, David (2003). The Common European Framework: principles, challenges, issues. *Neusprachliche Mitteilungen* 56/3 (2003), 130 – 140.
- Long, Michael H. (2007): *Problems in SLA*. London: Erlbaum.
- Lüdi, Georges (2003). Konzepte für mehrsprachige Repertoires als Zielvorstellungen für die Schule. Überlegungen zu einem „Gesamtsprachenkonzept“ für die Schulen in der Schweiz. In: G. Schneider / M. Clalüna (Hg.). *Mehr Sprache – mehrsprachig – mit Deutsch. Didaktische und politische Perspektiven*. München: iudicium, 276 – 291.
- Lüdi, Georges / Pekarek, Simona / Saudan, Victor (1999): *Apprentissage du français à l' intérieur et à l' extérieur de l' école*. Bern / Aarau: NFP (= PNR 33. Rapport de valorisation).
- Lüdi, Georges / Pekarek-Doehler, Simona / Saudan, Victor (2001): *Apprentissage du français à l'intérieur et à l'extérieur de l'école, rapport de valorisation, Berne/Aarau, PNR33/CSRE, 1999 (= PNR 33 – Efficacité de nos systèmes de formation)*.
- Lumley, Tom (2005): *Assessing Second Language Writing*. Frankfurt (Main): Peter Lang.
- Luoma, Sari (2004): *Assessing Speaking*. Cambridge: University Press.
- MacWhinney, Brian (1997): Second language acquisition and the Competition Model. In: Annette M. B. de Groot / Judith F. Kroll (Hg.): *Tutorials in Bilingualism: Psycholinguistic Perspectives*. Mahwah, NJ: Lawrence Erlbaum, 113-142.
- Maradan, Olivier (2004): *Stratégie de la CDIP et programme de travail pour la coordination de l'enseignement des langues à l'école obligatoire en Suisse*. In: *bulletin vals-asla* 79.
- Matthey, Marinette (1996): *Apprentissage d'une langue et interaction verbale. Sollicitation, transmission et construction de connaissances linguistiques en situation exolingue*. Bern u. a.: Lang (= Exploration: Recherches en sciences de l'éducation).
- McKay, Penny (2006): *Assessing Young Language Learners*. Cambridge: CUP (= Language Assessment).
- McNamara, Tim (2000): *Language Testing*. Oxford: University Press (= Oxford Introductions to Language Study).
- McNamara, Tim (1996): *Measuring second language performance*. Harlow: Longman.
- Merz-Grötsch, Jasmin (2000): *Schreiben als System, Bd 1: Schreibforschung und Schreibdidaktik. Ein Überblick*. Freiburg / Breisgau: Fillibach.
- Messick, S. (1989): *Validity*. In: Linn, R.L. (Hg.) *Educational measurement*. 3. Aufl. New York: American Council on Education/Macmillan, 13-103.
- Messick, S. (1994): *The interplay of evidence and consequences in the validation of performance assessments*. In: *Educational Researcher* 23(2), 13–23.
- Messick, S. (1996): *Validity and washback in language testing*. Princeton: ETS (Research Report).
- Milanovic, Michael (1996): *Guide for Examiners. Draft 1. (Modern Languages: Learning, Teaching, Assessment. A Common European Framework of reference)*. Strasbourg: Council of Europe (= Language Learning for European Citizenship CC-LANG (96)10 rev.). Online: <http://www.coe.int/T/DG4/Portfolio/documents/Guide%20October%202002%20revised%20version1.doc> (20.07.08).
- Mislevy, R.J. / Almond, R.G. / Lukas, J.F. (2004): *A brief introduction to evidence-centered design*. CSE Report 632. Los Angeles: National Center for Research on Evaluation.
- Mislevy, Robert J., Steinberg, Linda S. & Almond, Russell G. (2002): *Design and analysis in task-based language assessment*. In: *Language Testing*, 477-496.
- Molinaro, Agnese (2000): *Evaluation des compétences cognitives des élèves en situation d' apprentissage bilingue*. In: *L'école valdotaine. Cahier pédagogique* 46, 51-53.
- Morrow, K. (1979): *Communicative language testing: revolution or evolution?* In: Brumfit, C.J. and Johnson, K. (Hg.): *The communicative approach to language teaching*. Oxford: Oxford University Press, 143–57.

- Morrow, Keith (Hg.) (2004). Insights from the Common European Framework. Oxford: Oxford University Press.
- Moser, Urs (2000): Evaluation in den 'Fremdsprachen': Chancen und Grenzen. In: *Babylonia*. Themenheft: Lernerfahrungen reflektieren - Lernerfolge beurteilen. Réfléchir sur les expériences d'apprentissage - en évaluer les résultats 1, 10-14.
- Moser, Urs (2003): Klassenscockpit im Kanton Zürich Ergebnisse einer Befragung von Lehrerinnen und Lehrern der 6. Klassen über ihre Erfahrungen im Rahmen der Erprobung von Klassenscockpit im Schuljahr 2002/03. Bericht zuhanden der Bildungsdirektion des Kantons Zürich.
- Moser, Urs / Berweger, Simone (2003): Bildungsmonitoring Schweiz. Lehrplan und Leistungen. Thematischer Bericht der Erhebung PISA 2000. Neuchâtel: Bundesamt für Statistik (BFS) und Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK).
- Müller-Hartmann, A. & Schocker-v. Dittfurth, M. (Hrsg.) (2005). Aufgabenorientierung im Fremdsprachenunterricht. Task Based Language Learning and Teaching. Festschrift für M. Legutke. Tübingen: Narr.
- Nidegger, Christian (Hg.) (2003): A la fin de la 6e primaire. Enquête auprès de la volée 2000 des élèves de 6P. Acquis et compétences des élèves, représentations des parents et des enseignants. Genève: Service de la Recherche en Education.
- Nold, Günter (Hg.) (1992): Lernbedingungen und Lernstrategien. Welche Rolle spielen kognitive Verstehensstrukturen. Tübingen: Narr (= Tübinger Beiträge zur Linguistik 366).
- Nold, Günter (2003). DESI – a language assesment project in Germany and the pros and cons of large scale testing. In Claudia Finkbeiner / Sylvia Fehling (Hg.). Evaluation im Brennpunkt – Fremdsprachen lernen und lehren. Empirische Pädagogik 17, 3, 368-379.
- Nold, Günter / Rossa, Henning (2006): Leseverstehen. In: Beck, Bärbel / Klieme, Eckhard, Hrsg., *Sprachliche Kompetenzen: Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International)*. Weinheim und Basel: Beltz, 197-211.
- Norris, John M. (2002): Interpretations, intended uses and designs in task-based language assessment. In: *Language Testing*, 337-346.
- Norris, John M. / Brown, James D. / Hudson Thom / Yoshioka, Jim (1998): Designing second language performance assessments. Honolulu: University of Hawaii Press.
- North, Brian (2000): The development of a common framework scale of language proficiency. New York: Lang (= *Theoretical studies in second language acquisition* 8).
- North, Brian / Schneider, Günther (1998): Scaling, descriptors for language proficiency scales. In: *Language Testing* 2, 217-262.
- Nunan, David (1989): *Designing Tasks for the Communicative Classroom*. Cambridge: University Press.
- Nunan, David (2004): *Task-based Language Teaching*. Cambridge: CUP.
- ÖDaF - Mitteilungen. Themenheft: Was soll ich können müssen? Lernziele und Evaluationsverfahren für den DaF / DaZ-Unterricht. Wien: ÖDaF, 2003,1.
- OCDE (Éd.) (2003): Cadre d'évaluation de PISA 2003. En ligne : <http://213.253.134.43/oced/pdfs/browseit/9603052E.PDF>
- OCDE (2003): Reading Literacy. In: OECD, Hrsg., *The PISA 2003 Assessment Framework*, 107-129. Online: <http://www.pisa.oecd.org/dataoecd/38/52/33707212.pdf> (5.02.08)
- Oelkers, Jürgen / Reusser, Kurt (erscheint 2008): Qualität entwickeln, Standards sichern, mit Differenz umgehen. Expertise im Auftrag der vier Länder: Deutschland, Luxemburg, Österreich, Schweiz.
- O'Sullivan, Barry (2002): Learner acquaintanceship and oral proficiency test pair-task performance. In: *Language Testing* 19 / 3, 277-295.
- O'Sullivan, Barry / Weir, Cyril J. / Saville, Nick (2002): Using observation checklists to validate speaking-test tasks. In: *Language Testing* 19 / 1, 33-56.
- Phakiti, Aek (2007): *Strategic Competence and EFL Reading Test Performance: A Structural Equation Modeling Approach*. Frankfurt/M: Lang.
- Plaschy, Stephanie / Iwar Werlen (2000): Zweisprachige Schule Brig-Glis. Schlussbericht zum Projekt Classe Bilingue 1995-2000 an den Schulpräsidenten zuhanden des Gemeinderates der Stadtgemeinde Brig-Glis. Bern und Brig-Glis: Universitäres Forschungszentrum für Mehrsprachigkeit.
- Plaschy, Stephanie / Iwar Werlen (2001a): Schlussbericht über die Ergebnisse der Evaluation des ersten Jahrganges der zweisprachigen Matura. Zuhanden der Rektorate des Deutschen Gymnasiums Biel und des Gymnase Français Bienne. Bern und Brig-Glis: Universitäres Forschungszentrum für Mehrsprachigkeit.
- Plaschy, Stephanie / Iwar Werlen (2001b): Zweisprachige Matura Biel-Bienne Kurzbericht über die bisherigen Ergebnisse. Zuhanden der Rektorate des Deutschen Gymnasiums Biel und des Gymnase Français Bienne. Bern und Brig-Glis: Universitäres Forschungszentrum für Mehrsprachigkeit.
- Plaschy, Stephanie / Werlen, Iwar (2003a). Echange d'étudiants à la Haute Ecole Pédagogique due Valais. Rapport sur l'expérience et l'avis des étudiants, professeurs et praticiens-formateurs à la fin de la première année d'études. Brigue-Glis, Berne, Centre universitaires des recherche sur le plurilinguisme.

- Plaschy, Stephanie / Werlen, Iwar (2003b). Studierendenaustausch an der Pädagogischen Hochschule Wallis. Bericht über die Erfahrungen und Meinungen der beteiligten Studierenden, Dozierenden und Praktikumslehrpersonen am Ende des ersten Studienjahres. Brig-Glis, Bern, Universitäres Forschungszentrum für Mehrsprachigkeit.
- Pürschel, Heiner / Raatz, Ulrich (Hg.) (2001): Tests and Translation. Papers in memory of Christine Klein-Braley. Bochum: AKS (= FLF Fremdsprachen in Lehre und Forschung 27).
- Purpura, James E. (2000): Learner strategy use and performance on language tests: structural equation modeling Approach. Cambridge: University Press (= Studies in Language Testing 8).
- Quetz, Jürgen (2003): A1 - A2 - B1 - B2 - C1 - C2. Der Gemeinsame europäische Referenzrahmen. In: Deutsch als Fremdsprache. Zeitschrift zur Theorie und Praxis des Deutschunterrichts für Ausländer 1, 42-48.
- Rickheit, Gert / Herrmann, Theo / Deutsch, Werner (Hg.) (2003): Psycholinguistik - Psycholinguistics. Ein internationales Handbuch. Berlin: de Gruyter (=Handbücher zur Sprach- und Kommunikationswissenschaft, Band 24).
- Rickheit, Gert / Strohner, Hans (1993): Grundlagen der kognitiven Sprachverarbeitung. Tübingen: Francke.
- Riemer, Claudia (1997): Individuelle Unterschiede im Fremdsprachenerwerb. Eine Longitudinalstudie über die Wechselwirksamkeit ausgewählter Einflußfaktoren. Baltmannsweiler: Schneider Verlag Hohengehren.
- Riemer, Claudia (2000): Der interpretative Mehr-Methoden-Ansatz zur Untersuchung individueller Unterschiede“. In: Aguado, Karin (Hrsg.): Zur Methodologie in der empirischen Fremdsprachenforschung, Baltmannsweiler: Schneider Verlag Hohengehren, 93–104.
- Robinson, Peter (Hg.) (2001): Cognition and Second Language Instruction. Cambridge: University Press (= Cambridge Applied Linguistics).
- Robinson, Peter (2001): Task complexity, cognitive resources, and syllabus design: a triadic framework for examining task influences on SLA. In: Robinson, Peter (Hg.): Cognition and Second Language Instruction, 287-318.
- Robinson, Peter (2005): Cognitive complexity and task sequencing: Studies in a Componential Framework for second language task design. In: International Review of Applied Linguistics in Language Teaching (IRAL), 43, (1), 1-32.
- Rost, Michael (2002): Teaching and Researching Listening, Applied Linguistics in Action. Harlow: Pearson Education.
- Rupp, A. A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing*, 23, 441-474.
- Schneider, Günther (1999): Funktionen und wissenschaftliche Grundlagen der Kompetenzbeschreibungen im Sprachenportfolio. In: *Babylonia* 1, 29-33.
- Schneider, Günther (2000): Qualität zeigen. Fremdsprachenkenntnisse, Fremdsprachenlernen, Sprachenportfolio. In: Trier, Uri Peter (ed.): Bildungswirksamkeit zwischen Forschung und Politik. Nationales Forschungsprogramm 33, Wirksamkeit unserer Bildungssysteme. Chur, Zürich: Rüegger, S. 189-199.
- Schneider, Günther (2000) (Hg.): Themenheft Lernerfahrungen reflektieren - Lernerfolge beurteilen. Neue Tendenzen und Perspektiven im Bereich der Leistungsbeurteilung. *Babylonia* 1, 2000.
- Schneider, Günther (2001): Kompetenzbeschreibungen für das 'Europäische Sprachenportfolio'. In: Henrici, Gert / Zöfgen, Ekkehard (Hg.): Fremdsprachen lehren und lernen (FLuL) 30, 193-214.
- Schneider, Günther (2003): Der Europäische Referenzrahmen und die Mehrsprachigkeit. In: Neuner, Gerhard / Koithan, Ute (Hg.): Tagungsdokumentation 2003. 'Internationales Qualitätsnetz Deutsch als Fremdsprache'. Tagungsbeiträge und Arbeitsberichte zum Thema: 'Mehrsprachigkeit im Bereich Deutsch als Fremdsprache' 27.-29. Juni 2003, Universität Kassel, 87-108.
- Schneider, Günther (2005): Der "Gemeinsame europäische Referenzrahmen für Sprachen" als Grundlage von Bildungsstandards für die Fremdsprachen – Methodologische Probleme der Entwicklung und Adaptierung von Kompetenzbeschreibungen. In: Schweizerische Zeitschrift für Bildungswissenschaften 1, 13-37.
- Schneider, Günther / Lenz, Peter (2001): Portfolio européen des langues: Guide à l'usage des concepteurs. Strasbourg : Conseil de l'Europe.
- Schneider, Günther / Lenz, Peter (2001): Portfolio européen des langues: Guide à l' usage des concepteurs. Strasbourg: Conseil de l' Europe.
- Schneider, Günther / North, Brian (1999): 'In anderen Sprachen kann ich...!' Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit. Umsetzungsbericht. Bern: NFP (= NFP 33. Wirksamkeit unserer Bildungssysteme).
- Schneider, Günther / North, Brian (2000): Dans d' autres langues, je suis capable de...'. Échelles pour la description, l' évaluation et l' auto-évaluation des compétences en langues étrangères. Rapport de valorisation. Berne/Aarau: PNR 33/CSRE (= PNR 33. L'efficacité de nos systèmes de formation).
- Schneider, Günther / North, Brian (2000): Fremdsprachen können - was heisst das? Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit. Chur / Zürich: Rüegger (= NFP 33. Wirksamkeit unserer Bildungssysteme).
- Schneider, Günther / North, Brian / Koch, Leo / Schweizerische Konferenz der kantonalen Erziehungsdirektoren (Hg.) (2001): Portfolio européen des langues / Europäisches Sprachenportfolio / Portfolio europeo delle lingue / European Language Portfolio. Version pour jeunes et adultes. Version für Jugendliche und Erwachsene. Bern: BLMV.

- Schneider, Günther / Studer, Thomas (2007): Der ""Referenzrahmen"" und Bildungsstandards für die Fremdsprachen. In: Labudde, Peter (Hrsg.): Bildungsstandards am Gymnasium. Korsett oder Katalysator?, 219-228.
- Schwob, Irène (2002): Evaluation de l' enseignement / apprentissage bilingue en classe de 4e primaire de Monthey, Siere et Sion. Année scolaire 2000 / 2001. Rapport intermédiaire. Neuchâtel: IRDP.
- Serra, Cecilia (2002). Evaluation des zweisprachigen Unterrichts in Chur. Schulprojekt Zweisprachige Primarschule der Stadtschule Chur. Schuljahr 2001/2002. Zuhanden der Schuldirektion der Stadtschule Chur. Bern und Brig-Glis, Universitäres Forschungszentrum für Mehrsprachigkeit. Übersetzung aus dem Französischen: Stephanie Plaschy.
- Shaw, Stuart D. / Weir, Cyril (2007): Examining Writing. Cambridge: CUP.
- Shohamy, Elana (2003): The Power of Tests: A Critical Perspective on the Uses of Language Tests. Reviewed by Arieh Sherris. In: Chapelle, Carol A.: TESOL Quarterly. Teachers of English to speakers of other Languages 37 / 1, 189-190.
- Sigott, Günther (1993): Zur Lernbarkeit von Englisch und Französisch für deutsche Muttersprachler. Eine exploratorische Pilotstudie. Tübingen: Narr (= AAA - Buchreihe zu den Arbeiten aus Anglistik und Amerikanistik 6).
- Sigott, Günther (1998): Sprachkompetenzmessung mit dem C-Test: Ein Resumé. In: ÖDaF - Mitteilungen. Themenheft: Testen, Prüfen, Leistungsbeurteilung und Evaluation. LernerInnen_Autonomie 2, 9-20.
- Sigott, Guenther/ Gassner, Otmar/ Mewald, Claudia/ Siller, Klaus (2007): *E8-Standardstests. Entwicklung der Tests für die rezeptiven Fertigkeiten: Überblick*. Wien: Bundesministerium für Bildung, Wissenschaft und Kultur. (=LTC Technical Report 1)
- Skehan, P. & Foster, P. (2001). Cognition and tasks. In P. Robinson (Hrsg.). Cognition and Second Language Instruction. Cambridge: University Press, 183-205.
- Skehan, Peter (1998): A Cognitive Approach to Language Learning. Oxford: University Press.
- Skehan, Peter / Foster, Pauline (2001): Cognition and tasks. In: Robinson, Peter (Hg.): Cognition and Second Language Instruction, 183-205.
- Spolsky, Bernard (o. J.): Assessment and testing: notes for the third millennium. Online: <http://www.russnet.org/fipse/SpolskyWhitePaper.html> (30.05.08)
- Stede, Manfred (2007): *Korpusgestützte Textanalyse. Grundzüge der ebenenorientierten Textlinguistik*. Tübingen: Gunter Narr Verlag.
- Steinberg, Danny D. / Nagata, Hiroshia / Aline, David P. (2001): Psycholinguistics: Language, Mind and World. Harlow: Pearson Education.
- Stotz, Daniel (2000): Evaluation im frühen teilimmersiven Fremdsprachenunterricht. In: Babylonia. Themenheft: Lernerfahrungen reflektieren - Lernerfolge beurteilen. Réfléchir sur les expériences d' apprentissage - en évaluer les résultats 1, 19-22.
- Studer, Thomas (2000): Internationale Sprachenzertifikate aus der Sicht von Schweizer TestautorInnen. Aspekte der Testentstehung am Beispiel des Zertifikats Deutsch, der Zentralen Mittelstufenprüfung und der DELF/DALF-Diplome In: Babylonia. Themenheft: Lernerfahrungen reflektieren - Lernerfolge beurteilen. Réfléchir sur les expériences d' apprentissage - en évaluer les résultats 1, 62-66.
- Studer, Thomas (2004): Internationale Zertifikate für Deutsch als Fremdsprache in der Schweiz. In: bulletin vals-asla 79.
- Studer, Thomas / Mettler, Monika / Lenz, Peter (2004). Das Projekt IEF: Instrumente für die Evaluation von Fremdsprachenkompetenzen (Englisch/Französisch). Schweizerische Zeitschrift für Bildungswissenschaften, 419-434.
- Studer, Thomas u. a. (1999): Zertifikat Deutsch. Lernziele und Testformat. Frankfurt a.M. WBT/ ÖSD/ Goethe-Institut/ EDK.
- Tschirmer, Erwin (2001): Die Evaluation fremdsprachlicher mündlicher Handlungskompetenz: Ein Problemaufriss. In: Henrici, Gert / Zöfgen, Ekkehard (Hg.): Fremdsprachen lehren und lernen (FLuL) 30, 87-115.
- Tunger, Verena / Werlen, Iwar (2003a). Échange intersite à la Haute École Pédagogique du Valais. Rapport sur l'expérience et l'avis des étudiants, professeurs et praticiens-formateurs à la fin de la première et de la deuxième année d'études. Brigue-Glis, Berne, Centre universitaires des recherche sur le plurilinguisme.
- Tunger, Verena / Werlen, Iwar (2003b). Studierendenaustausch an der Pädagogischen Hochschule Wallis. Bericht über die Erfahrungen und Meinungen der beteiligten Studierenden, Dozierenden und Praktikumslehrspersonen am Ende des ersten und des zweiten Studienjahres. Brig-Glis, Bern, Universitäres Forschungszentrum für Mehrsprachigkeit.
- Urquhart, Sandy A.H. / Weir, Cyril J. (1998): Reading in a second language: process, product and practice. London / New York: Longman (= Applied Linguistics and Language Study).
- Veltcheff, Caroline / Hilton, Stanley (2003): Evaluation en FLE. Paris: Hachette.
- Vollmer, Hellmut. J. (2003). Ein gemeinsamer europäischer Referenzrahmen für Sprachen: Nicht mehr, nicht weniger In Bausch / Christ / Königs / Krumm (Hg.). Der Gemeinsame europäische Referenzrahmen in der Diskussion. Tübingen: Narr, 192-206.
- Weidinger, Walter et. al. (2003): Erprobung und Evaluierung von Bildungsstandards an Wiener Pflichtschulen der Sekundarstufe I. Studie Bildungsstandards. Projektinformaton Mai 2003. Wien: Stadtschulrat.

- Weinert, F. E. (2001): Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In: F. E. Weinert (Hrsg.): Leistungsmessungen in Schulen. Weinheim und Basel: Beltz Verlag, S. 17–31.
- Weir, Cyril J. (1993): Understanding and Developing Language Tests. London: Prentice Hall.
- Weir, Cyril J. (2005a): Language Testing and Validation: An Evidence-Based Approach. Basingstoke: Palgrave/Macmillan.
- Weir, Cyril J. (2005b): Limitations of the Common European framework for developing comparable examinations and tests. In: Language Testing, 3, 281-300.
- Weiss, Jacques (1994): Evaluation interne de l' expérience d' apprentissage bilingue de Sierre. Rapport à l' intention du groupe de coordination. Neuchâtel: IRDP (= IRDP Recherches).
- Weiss, Jacques (1998): Une évaluation informative pour la régulation des apprentissages et des formations. Neuenburg: IRDP (= IRDP Recherches).
- Werlen, Iwar (2003). Flüssiges Sprechen vortäuschen: Wiederholungen als Technik zur Erzeugung flüssigen Sprechens bei Sprecherinnen von Deutsch als L 2. In: Mondada, Lorenza / Pekarek Doehler, Simona (éds.), Plurilinguisme Mehrsprachigkeit Plurilingualism. Enjeux identitaires, socio-culturels et éducatifs. Festschrift pour Georges Lüdi. Tübingen, Basel, Francke, 179-195.
- Werlen, Iwar / Fuchs, Gabriela (1999): Zweisprachige Klassen der Bieler Gymnasien, Projektverlauf 1998/99. Befragung der Schülerinnen und Schüler, der Eltern und der Lehrerschaft. Brig-Glis/ Bern: Universitäres Forschungszentrum für Mehrsprachigkeit.
- Widdowson, H.G. (1989): Knowledge of language and ability for use. In: Applied Linguistics, 10 (2), 128-137.
- Wilson, Mark (2004): Constructing measures. An item response modeling approach. Mahwah, N.J.: Erlbaum.
- Wolfe-Quintero, Kate / Inagaki, Shunji / Kim, Hae-Young (1998): Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity. Honolulu: University Press.
- Wolff, Dieter (2002): *Fremdsprachenlernen als Konstruktion. Grundlagen für eine konstruktivistische Fremdsprachendidaktik*. Frankfurt/Main etc.: Lang.
- Yuan, Fangyuan / Ellis, Rod (2003): The Effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. In: Applied Linguistics 24 / 1, 1-27.
- Zarate, Geneviève / Gohard-Radenkovic, Aline / Lussier, Denise / Penz, Hermine (2003): Médiation culturelle et didactique des langues. Strasbourg: European Centre for Modern Languages / Conseil de l'Europe.
- Ziberi-Luginbühl, Johanna (1999): Zweitsprachunterricht im obligatorischen Schulsystem. Bern: NFP (= NFP 33. Umsetzungsbericht).
- Zimmermann, Hansmartin / Iwar Werlen (1996). Das Projekt "Zweitsprachunterricht im obligatorischen Schulsystem": Konzept, empirische Daten und erste Ergebnisse. In: : Anne-Claude Berthoud (éd), Acquisition des compétences discursives dans un contexte plurilingue. Bulletin Suisse de linguistique appliquée (VALS/ASLA) 64, octobre, 35-59.
- Zydati, Wolfgang (2002): Leistungsentwicklung und Sprachstandserhebungen im Englischunterricht. Methoden und Ergebnisse der Evaluierung eines Schulversuchs zur Begabtenförderung: Gymnasiale Regel- und Expressklassen im Vergleich. Frankfurt/M. u. a.: Lang (= KFU - Kolloquium Fremdsprachenunterricht 8).
- Zydati, Wolfgang (2005): Bildungsstandards und Kompetenzniveau im Englischunterricht. Konzepte, Empirie, Kritik und Konsequenzen. Frankfurt u. a.: Lang (= Kolloquium Fremdsprachenunterricht 22).

Logiciels statistiques:

- ConQuest Generalised Item Response Modelling Software, Build Jan. 10, 2006. Wu, Margaret / Adams, Ray / Wilson, Mark / Haldane, Sam. Australian Council for Educational Research; University of California, Berkeley.
- Facets, v. 3.63 (2007). Linacre, John Mike. Chicago.
- JMP - The Statistical Discovery Software. SAS Institute Inc. Cary, N.C.
- R. The R Project for Statistical Computing. Online: <http://www.r-project.org>.
- SPSS, v. 16 (2007). SPSS, Inc. Chicago.
- Winsteps, v. 3.64 (2007). Linacre, John Mike. Chicago.