ÜGK/COFO Mathematics 2016
Audit Report

Dr Antoine Fischbach & Dr Sonja Ugen

Commissioned by EDK/CDIP

Upon the Request of KOSTA HarmoS

Luxembourg, 23 February 2018

The auditors would like to thank everyone involved for preparing and implementing the ÜGK/COFO mathematics 2016 audit, for making the documentation available, and for participating in the hearings.

Table of Contents

I. Introduction 4

1. Context and Contractual Mission 5

2. A Word about the Auditors 6

3. Auditing Procedure 8

II. Results of the Audit 11

1. Item and Test Development 13

2. Test Administration 16

3. Data Processing and Analysis 17

4. Standard-Setting 20

5. Measurement of Minimum Standards 21

III. Summary and Recommendations 24

1. Summary 24

2. Recommendations 25

Index of Acronyms and Abbreviations 32

Appendix A – Stakeholder Hearings 34

Appendix B – Documentation 35

I. Introduction

This report presents the findings of the audit of the 2016 Swiss large-scale assessments of basic mathematics competences (ÜGK/COFO) conducted between November 2017 and February 2018. The audit was commissioned by the Swiss Conference of Cantonal Ministers of Education (EDK/CDIP) upon the request of the coordinating body of the Intercantonal Agreement on Harmonisation of Compulsory Education (KOSTA HarmoS).

Before presenting the contents of the report, the auditors would like to emphasise that the ÜGK/COFO audit was conducted under ideal conditions. The documentation that was provided was exhaustive and delivered on time. Without exception, the auditors' exchanges with the stakeholders were cooperative and constructive. Above all, the auditors are under the impression that the audit process was fully supported and met with transparency by all stakeholders involved.

The present report was composed to the best of the auditors' knowledge and was based on the data that were collected as described in the *Auditing Procedure* section. The auditors hereby explicitly declare that they were free from any conflict of interest during the evaluation and that they have no personal or professional stakes in the ÜGK/COFO operation.

For the sake of clarity, the auditors decided not to translate official names and abbreviations/acronyms into English (see also *Index of Acronyms and Abbreviations*). If available, German and French terms are used jointly to maximise intelligibility.

The report from the audit consists of three parts: (I) The *Introduction* section establishes the setting and the methodology applied in the evaluation. (II) The *Results of the Audit* section describes the main findings concerning the ÜGK/COFO processes with an emphasis on the contracting body's questions (see *Context and Contractual Mission*). (III) Finally, in the *Summary and Recommendations* section, the auditors lay out some recommendations for future ÜGK/COFO operations on the basis of their observations and conclusions.

In the following, the auditors will briefly recapitulate the context and the mission of the audit as stated in the contract, present the background of the two responsible auditors, and outline the applied audit procedure.

1. Context and Contractual Mission

Upon the request of the EDK/CDIP, in 2016, Switzerland administered the first national assessments of basic mathematics competences (ÜGK/COFO) for students in Grade 11. On the basis of the educational objectives published in 2011 by the 26 Cantonal Ministers of Education, the 2016 ÜGK/COFO aimed to verify the achievement and harmonisation of basic competences (Grundkompetenzen; compétences fondamentales) in mathematics that the vast majority of pupils were expected to master by the end of the compulsory school curriculum in Switzerland. Didactics and teaching representatives of the three linguistic regions developed these educational objectives on the basis of the existing cantonal curricula. The resulting HarmoS minimum standards[1] have since been integrated into the new regional curricula (i.e., Lehrplan 21, Plan d'études romand, Piano di studio).

On the basis of the theoretical descriptors in the HarmoS reference documentation, mathematics didactics experts, secondary school teachers, and psychometricians elaborated the test items (tasks) for the mathematics data collection in 2016. The resulting mathematics test was pretested in 2015. The main test in 2016 included 180 items that were implemented in a cluster rotation design. After the main data collection, cut-off scores for the HarmoS minimum standards (Grundkompetenzen; compétences fondamentales) were determined during a two-day workshop in which experts of didactics and teaching participated in a standard-setting process (Schwellenwertsetzung).

KOSTA HarmoS, the strategic body of the ÜGK/COFO, decided to request an audit report on the processes of ÜGK/COFO item development and standard-setting.

[1] "Basic competences expected to be mastered by the vast majority of pupils" is the textbook definition of "minimum standards" sensu Klieme. This is why the auditors henceforth adopt this more common terminology.

As stated in the contract, the auditors were mandated to provide answers to the three following questions:

(1) Does the process of the 2016 ÜGK/COFO mathematics item and test development meet scientific standards?

(2) Does the process of the 2016 ÜGK/COFO mathematics standard-setting meet scientific standards?

(3) Do the 2016 ÜGK/COFO mathematics items assess the HarmoS minimum standards (Grundkompetenzen; compétences fondamentales) released by the EDK/CDIP regarding (a) the competence domains and aspects as well as (b) the difficulty?

2. A Word about the Auditors

Dr Antoine Fischbach[2] and Dr Sonja Ugen[3], referred to as "the auditors", currently form the managing directorate of the Luxembourg Centre for Educational Testing (LUCET)—a 30-person interdisciplinary research and transfer structure at the University of Luxembourg—which they co-created and executively managed from 2012 to 2015. The LUCET's raison d'être and most prominent research undertaking is the implementation, enhancement, and assurance of the Luxembourg school monitoring programme (Épreuves Standardisées; ÉpStan), which is aimed at facilitating evidence-based decision making in national education. Closely related to the latter is the setup of a unique longitudinal database in which panels are entire cohorts. The database contains information about the evolution of students' competency profiles and their educational pathways. Additional mission-oriented LUCET research projects include national analyses and reports on international large-scale studies (e.g., PISA), (large-scale) cognitive and language testing, university admissions testing, and student course evaluations.

[2] Dr Antoine Fischbach holds a degree in educational sciences, an MSc "evaluation and assessment" in psychology (both from the University of Luxembourg), and a doctor rerum naturalium in psychology (University of Trier).

[3] Dr Sonja Ugen holds a BSc in psychology (London Guildhall University), an MSc in developmental psychology (Lancaster University), a DEA in psychological sciences, and a doctorate in psychological and educational sciences (both from the Université libre de Bruxelles).

Crucial for the present audit, the auditors were able to draw on the hands-on experience they gained from each part of the ÉpStan operation, ranging from item development to reporting, and they designed most of the processes still in place today. Moreover, the auditors have hands-on experience with numerous international large-scale studies including PISA, PIRLS, ICCS, and HBSC. In many of the aforementioned national assessments, many of which are also high-stakes assessments, the LUCET's in-house online assessment system OASYS is used. Thus, given that many of these assessments are completely computer- and web-based, the auditors have been able to build solid expertise in managing ambitious and complex technology-rich research endeavours. Note that the auditors were also actively involved in the original TAO initiative.

The auditors hold several high-level administrative and scientific positions, both inside (e.g., Faculty Council, University Council, Doctoral School in Humanities and Social Sciences Governing Board, University of Luxembourg Leadership Academy Governing Board) and outside of their institution (e.g., Luxembourg High Council for Education, ADMEE Europe Governing Board, Scientific Committee of French for Business at CCI Paris Île-de-France). Important for the present evaluation, through their long-term memberships in the inter-institutional BELDACH (formerly DACHL) network, the auditors can look back on almost a decade of high-level insights into the landscape of Swiss education.

Over the years, the auditors have (co)secured over 5M € worth of (mainly mission-oriented) research funding, (co)authored numerous peer-reviewed scientific papers (in 3 different languages), accumulated a considerable record of refereed conference contributions and (invited) talks (in 4 different languages), and served on a regular basis as ad hoc reviewers for leading journals in the field. The auditors have (co)authored and (co)edited essential study reports on national education and have been consulted by national (e.g., Luxembourg Ministry of Education, Luxembourg High Council for Sustainable Development) as well as international stakeholders (e.g., European Commission) to address questions of educational assessment and educational quality management. The

auditors hold full PhD supervision rights and currently (co)supervise 9 PhD theses revolving around methodological and technological challenges in (large-scale) assessment in highly diverse and multilingual learning environments.

In 2016, the LUCET underwent an extensive external research evaluation by the Swiss-based INTERFACE GmbH. In their final evaluation report, the experts stated that they assessed the LUCET's performance as "outstanding and comparable to excellent educational research facilities in leading countries". Given the overall success of the LUCET, the auditors are currently coordinating and implementing a merger between the LUCET and the Luxembourg Centre for School Development.

3. Auditing Procedure

The main steps of the auditing procedure involved the transfer of documentation, a face-to-face meeting with EDK/CDIP stakeholders to introduce the ÜGK/COFO operation, an in-depth analysis of the documentation provided, a series of hearings with the collaborators involved in the main ÜGK/COFO processes, and the drafting and transmission of the final report. Even though a self-evaluation report prepared by the auditees is often implemented as the starting point of an audit process, it was not possible to integrate such a report here due to the short timeframe.

Concretely, after a brief exchange with the ÜGK/COFO project coordinator, the auditors officially accepted the mandate on 28 November 2017. Subsequently, an exhaustive list of documentation was provided to the auditors (7 December). Soon thereafter (13 December), a first face-to-face meeting with EDK/CDIP stakeholders (ÜGK/COFO project coordination and ADB/BDT team) took place at the University of Luxembourg's Belval Campus in Esch-sur-Alzette. The auditees presented the ÜGK/COFO operation and handed over additional documentation. For their part, the auditors presented the audit procedures, as described above, to the stakeholders. Over the course of the following weeks, the auditors studied the documentation thoroughly and drew up a list of key ÜGK/COFO stakeholders whom they wanted to interview to complement the

picture drawn by the documentation. Between 23 and 29 January 2018, nine semi-standardised hearings were organised with stakeholders of all the key steps of the ÜGK/COFO operation (i.e., development of the basic competences for mathematics, item development, standard-setting procedure, coding of (half)open items, scientific consortium, data collection, item database, and project coordination). The hearings were organised by the ÜGK/COFO project coordination team and took place either face-to-face in the EDK headquarters in Bern or online via video conferencing. The interview guidelines included a common core of questions that were identical for all interviewees in a first step and included more specific questions related to the area of expertise of the interviewee in a second step. The hearings were structured so that at least two interviewees responded to each specific question. Participants could respond in the language of their choice. Prior to the hearing, participants were informed that all information they provided would be handled confidentially. Each hearing lasted for approximately one hour. *Appendix A – Stakeholder Hearings* lists all the participants who were present at the hearings. The hearings were very well organised, and all invited participants were present as planned. Some participants provided additional documentation prior to or after the hearing. *Appendix B – Documentation* lists the complete materials the auditors received throughout the auditing process to which the auditors added the educational curricula of two linguistic regions (i.e., Lehrplan 21, Plan d'études romand). Please note that no other documentation sources were considered in the auditing process.

Following the on-site interviews on 25 January 2018, the auditors briefly discussed the first tentative hypotheses and conclusions with the ÜGK/COFO project coordination team to ensure factual comprehension and to avoid possible misunderstandings. After the hearings, the auditors focused on the report. Note that the auditors had mathematics didactics as well as psychometrics experts at their disposal to answer questions about details. It goes without saying that the confidentiality terms of the audit contract were strictly respected at all times. On 19 February 2018, the auditors submitted

a draft report to the ÜGK/COFO coordination team for fact-checking. The ÜGK/COFO coordination team was allowed to rectify stated facts until 22 February. Comments about the outcome or the conclusions of the audit were not considered by the auditors. Note that the dates for fact-checking had already been arranged and communicated in December 2017. The final audit report was submitted to the EDK/CDIP before 28 February 2018 as agreed upon in the contract.

Upon request, the auditors will present the audit outcome and recommendations—in German and French with bilingual visual support—to the EDK/CDIP.

II. Results of the Audit

In the results section of the present report, one chapter is dedicated to each question specified in the *Context and Contractual Mission* section: *Item and Test Development*, *Standard-Setting*, and the *Measurement of Minimum Standards* (Grundkompetenzen; compétences fondamentales). The auditors added two chapters on *Test Administration* and *Data Processing and Analysis* as these processes were integral parts of the ÜGK/COFO operation and could possibly have impacted the three aspects under investigation. Two project aspects are however not discussed in detail: data sampling and IT issues because addressing these issues would have gone beyond the scope of the current audit. For the data sampling, the procedures that were applied were similar to those of internationally recognised studies (e.g., PISA), and the auditors considered these to be adequate. Regarding IT, the auditors refer to technological problems that occurred, but they discuss the impact of these problems at only the level of the general project.

Before analysing each part of the operation in detail, some key observations are reported. First, the auditors would like to emphasise that both HarmoS and ÜGK/COFO are extremely ambitious endeavours bearing high stakes that extend far beyond educational matters. As evidenced by the exhaustive meeting notes, papers, and reports, the undertaking was meticulously planned, and a great deal of thought was put into it (see *Appendix B – Documentation*; especially the 2004 HarmoS white book and the 2017 handbook on large-scale assessments in Switzerland). The HarmoS concept and more specifically the ÜGK/COFO provide the opportunity for politics and science to reclaim the educational assessment field, which has been seized by diverse commercial actors in Switzerland in recent years.

The elaborate ÜGK/COFO organisational chart provides additional evidence of how well-planned the general endeavour is. The ÜGK/COFO even encompasses the FORS centre to handle and disseminate the (future) ÜGK/COFO database(s) for research purposes. However, not only will the ÜGK/COFO database be relevant for research purposes but it will also—and above all—

constitute a future asset for educational politics and evidence-based policymaking.

The complexity of the ÜGK/COFO organisational chart with many different bodies in different geographical locations albeit being exhaustive—even if not all implemented at this stage—also carries the challenge of the coordination of the project and the people involved while taking personal, linguistic and cultural differences into account. The auditors have the impression that the current ÜGK/COFO management situation is suboptimal because the project coordination and the contracting body are not separate entities and are thus not able to independently carry out their respective functions. Whereas many of the bodies have official contracts that state their missions, a number of actors within the ÜGK/COFO have provided crucial input into the operation without an explicit mandate or a clear mission, thus causing frustration due to the lack of official recognition and blurred areas of command. The auditors would like to emphasise, however, that the actions of all the parties involved were motivated by good intentions and were based on the best knowledge available. The reason the entire ÜGK/COFO operation could be successfully carried out was that so many parties efficiently fulfilled their part of the job even under severe time constraints. During the operation, however, some tensions and conflict situations arose between different actors, but the actors were able to transform these conflicts into acceptable and professional working situations towards the end. Such situations are not surprising considering the high stakes of the project along with the omnipresent time pressure.

Although the ÜGK/COFO initiative in itself was well planned in advance, the 2016 implementation in mathematics appeared to be rushed. Indeed, the ÜGK/COFO in mathematics involved a steep learning curve in the area of national large-scale assessments. The expertise and infrastructure were quickly built and were applied in parallel. The tight time frame put even more pressure on the different parties concerned when some technological issues slowed down parts of the operation.

The auditors were able to determine that the political governance of the ÜGK/COFO operation is reliable and well-functioning. However, the corresponding scientific governance (WiKo) is currently not yet in place. Sound scientific guidance is of utmost necessity for high-stakes large-scale assessments such as the ÜGK/COFO. The setup and active functioning of the WiKo should be considered a priority.

The subsequent paragraphs follow the natural order of the steps involved in the elaboration of an assessment tool beginning with the content (*Item and Test Development*) and ending with the targeted measurements (*Measurement of Minimum Standards*).

1. Item and Test Development

(1) Item development. Item development is an extremely important part of the process of developing large-scale assessments. Indeed, the items form the basis of the data that are analysed and interpreted, upon which critical educational decisions follow.

For the 2016 ÜGK/COFO in mathematics, a substantial number of items were created (i.e., 180), but only 132 could be retained for the final statistical model. In terms of content, the items can be related to the competence domains and aspects as described in the documentation on the theoretical references. Although some items could be sorted into multiple domains and/or aspects, and the classification rationale was not always obvious, the developed items seem to correspond to the underlying basic competence model (for more details, see *Measurement of Minimum Standards*).

In general, the chosen item formats—mainly multiple-choice and (half)open text—are appropriate for large-scale assessment applications. Concerning the multiple-choice items, however, the response formats were not always applied consistently (e.g., 1 out of 4 vs. 1 out of 7). This inconsistency thus artificially impacted item difficulty—not through item content but through the use of easier or more difficult response formats with different cognitive and/or memory demands, as well as different guessing probabilities, which, in return, might impact test validity. The distractors (i.e., the wrong answers in multiple-choice items) were rather well-designed in terms of content and visual appearance. Developing equally

attractive distractors is far from trivial and key to high-quality closed-format items. Accordingly, future item development processes should include systematic distractor analyses that are based on empirical (pretest) data. For (half)open text items, the coding guidelines were initially missing. As a matter of fact, the guidelines had to be developed post hoc under the guidance of an expert in mathematics didactics who was not part of the item development group. Written coding guides are an indispensable part of the item development process in which the expected responses are explicitly defined. A subjective interpretation of responses can impact item difficulty and is not suitable for large-scale tests because inter-rater reliability cannot be guaranteed. The auditors would like to stress that the post hoc coding guidelines and the coding process were managed professionally even under severe time constraints and with limited technical means (for more details, see *Data Processing and Analysis*).

The ÜGK/COFO mathematics items also involved a great deal of text. It is well known that language proficiency has an important impact on mathematics performance. For assessments in mathematics, it is thus recommended that the language load be reduced as much as possible. Especially with the HarmoS objective in mind, it is of utmost importance to pay attention to aspects of linguistics when developing items in three different languages and to reduce the influence of language on performance as much as possible. An adequate and coherent use of illustrations can help to reduce the language component in mathematics items. In the ÜGK/COFO, however, item illustrations were partly inappropriate, confusing, or only decorative in nature, which did not help to convey item content.

Overall, the *exact* item development process remains unclear to the auditors. The auditors were not provided with proper documentation describing the ÜGK/COFO item development process, and even throughout the hearings, the auditors did not gain satisfactory insights into the item development modus operandi. There seems to be a lack of standardised guidelines and procedures. Item developers did not receive any kind of methodological training

prior to development. Even though it remains somewhat unclear who *exactly* participated in the item development process, the item developers were most likely experts in didactics but with only a little expertise in item and test development as the numerous (half)open items and diverse closed-format responses demonstrate. Ideally, an item development group consists of a balanced number of didactics experts, practitioners, and psychometricians to ensure the items are theoretically well-founded, feasible, presented in an appropriate format, and properly pretested. To make things even more complicated, the ÜGK/COFO item development was marked by interpersonal and intercultural (didactical) differences that resulted in a clash of very different working styles and ultimately in a (short-lived) conflict situation. More transparent management with clear decision processes and efficient coordination would most likely have helped the developers avoid the disagreements that occurred during the ÜGK/COFO item development. In the end, however, the various actors managed to deal constructively with the—certainly challenging—situation, and the item development process could be finalised.

In general, in the ÜGK/COFO operation, the item development part has been underestimated in terms of time, resources, and expertise and has thus been understaffed and underfinanced. Crucial processes (e.g., the translation processes) have not been formalised, and their workloads have been underestimated. In addition, the available resources often had to be used to address technological issues concerning the database. Thus, valuable time and resources were not available for focusing on the essential parts of the mission (i.e., item content).

(2) Test development. The documentation on the test development of the ÜGK/COFO was excellent. The test designs and frameworks were well balanced, and the applied cluster rotation met high scientific standards. Unfortunately, however, the items that were included in the main test were never adequately pretested; half were adjusted after a first pretest, and half were newly developed afterwards. The auditors are aware that this was mainly due to the severe time constraints. Nevertheless, the auditors note that instead

of developing an extensive number of items, more time should be invested in developing fewer high-quality items, even if this involves a less complex multimatrix design, especially as the outcome of the standard-setting procedure was the use of a single scale. In all fairness, the auditors are well aware that this single-scale outcome was not known when the items were being developed and the test was being designed.

Despite substantial shortcomings, the final results (i.e., the 132 retained mathematics items) are acceptable. All items are theoretically well-grounded, and the translations are not perfect, but they are acceptable. Moreover, all items are locally independent from one another—a prerequisite for the chosen data analytic framework (see *Data Processing and Analysis*). The resulting student competence estimations are normally distributed, and the various plausibility analyses (e.g., the comparison of cantonal distributions or proportions of low achievers) are convincing, leading the auditors to the conclusion that the 2016 ÜGK/COFO data are plausible and interpretable. To conclude the present section, the auditors would like to underscore once more that, even under the aforementioned precarious conditions, the item and test development process was able to succeed overall thanks to the exemplary engagement of various—and not necessarily explicitly mandated—actors.

2. Test Administration

In the eyes of the auditors, the ÜGK/COFO administration was highly professional and conducted in an exemplary fashion. The regional test administration centres benefit from considerable experience in carrying out international large-scale assessments (e.g., PISA) and consequently devised comparably stringent scientific procedures for the ÜGK/COFO in mathematics. The large set of manuals that was provided was available in the three main regional languages and included comprehensive and extensive information about technical, practical, as well as conceptual aspects of the ÜGK/COFO endeavour. The auditors would like to explicitly point out the very high level of quality and thoroughness reflected in the trilingual documentation. Prior to the field tests, the test administrators attended a training program. Remarkably, systematic quality

assurance and quality control processes (e.g., the monitoring of test sessions) were included in the overall procedure to guarantee optimal assessment conditions. Significant for the present audit, the test sessions were observed to convey the high levels of motivation shown by the participating students and a high level of acceptance of the ÜGK/COFO amongst the teachers and schools. This motivational aspect represents an additional argument for the plausibility of the 2016 ÜGK/COFO data. In particular, on a conceptual level, the acceptance of the national ÜGK/COFO assessment seems to be generally higher than for other international large-scale assessments. The observed/reported high level of acceptance of the ÜGK/COFO represents a non-negligible asset. It is also a good starting point for the HarmoS endeavour in general, for the successful implementation and application of the new curricula, and for fruitful future cooperation with practitioners and schools.

3. Data Processing and Analysis

(1) Data processing. As already noted (see *Item and Test Development*), coding guidelines for (half)open test items, which are an integral part of item development in the eyes of the auditors, were not provided by the item developers as the auditors would have expected. Consequently, the entities responsible for the coding of (half)open items had to reverse-engineer the guidelines before their actual work could begin. The coding itself was supposed to take place on an online platform, but its setup failed due to web-hosting issues. Eventually, a less automated, more cumbersome, and relatively inefficient and more error-prone PDF-based solution was achieved. Both aforementioned aspects (i.e., the nonexistent guidelines and the suboptimal technological solution) served only to augment the already severe time pressure. To make ends meet, quality assurance processes had to be cut back, and the planned inter-rater reliability check—an explicit desideratum for future ÜGK/COFO operations—was sacrificed. However, the training of coders took place as planned, and a coder query was organised throughout the entire process. Given the initial lack of coding guidelines, this second aspect was particularly important for data

quality. To sum up, the solid expertise acquired from administering previous international large-scale assessments, intelligent crisis management, and high level of flexibility eventually circumvented a looming disaster in the handling of (half)open test items.

(2) Data analysis. Regarding the data analysis, the documentation was excellent and very comprehensible. All analyses were executed in the powerful open-source R statistical environment, a versatile and sustainable technological choice. The ÜGK/COFO sample size was sufficient for the chosen IRT data analytic framework, the amount of missing and/or invalid data was well within the boundaries of an acceptable level, and the handling of missing/invalid data followed standard operating procedures. Rasch is a solid and safe choice for the scaling of competence data. Given that there is not yet a trend in competence over time requiring protection, it would alternatively have been possible to opt for the two-parameter Birnbaum model, which has gained popularity in international large-scale assessments due to its slightly better discrimination over the one-parameter Rasch model. That said, there are clear pros and cons for both choices, and there is no single best solution. When evaluating model fit, the ÜGK/COFO data analysts relied on three criteria: (a) item discrimination, (b) item infit, and (c) visual analysis of the ICCs. For (a) and (b), the chosen cut-offs were well selected and rather strict in international comparison. Regarding (c), ICCs are important model fit indicators that are all too often neglected as a source of information; systematically screening the ICCs was definitely a strength in the ÜGK/COFO data analytic procedure. DIF was investigated by means of logistic regression, which is a scientifically well-established method. The DIF grouping variables were satisfactory and theoretically well-grounded. For the DIF analysis, students' proficiency was operationalised as WLE scores, and again, this was a good choice because Warm's estimates are widely considered to be the best person parameter estimates. The magnitude of DIF was classified along the widely used ETS categories. However, only test items of ETS Type "c" (i.e., items with large DIF) were flagged for in-depth analysis and possible elimination. In the eyes of the auditors,

this focus on Type “c” items only is a noteworthy flaw in the ÜGK/COFO data analysis. The chosen DIF cut-off is at the very upper end of DIF generosity, and a less lenient cut-off should be considered for future investigations.

Notably, only 132 out of 180 items passed all quality control steps and remained in the final scaling model. Over one quarter of item loss is very high and too high for a main test operation. Critically, and as previously suggested, with a stricter DIF cut-off, the item loss would have been even higher. The emphasised item loss is not a matter of analysis but is again a matter of suboptimal item development and, first and foremost, insufficient pretesting. It is the understanding of the auditors that, primarily out of time constraints, about half of all the 2016 ÜGK/COFO mathematics items were never pretested, and the other half was substantially adjusted after the pretest. Neglecting item pretesting is a false economy that led in this specific case to a 25% loss in information and 25% in investments without return. Hence, future ÜGK/COFO operations should imperatively treat pretests with due diligence.

The applied plausible values technique, a procedure for correcting relationships between latent variables and covariates, is the adequate choice for system reporting of large-scale assessment data, as alternative procedures for estimating person parameters result in biased estimates of relationships at the population level. The set of covariates incorporated in the drawing of plausible values is satisfactory and theoretically well-grounded. The naturally imposed hierarchical structure of the ÜGK/COFO data was duly respected at all steps of the operation, and the imputation process of covariates and plausible values alike can be considered scientific best practice. Finally, the selected response probability for the minimum standard cut-off followed international conventions.

All in all, the ÜGK/COFO data processing and analysis was impressive and fully in line with the current state-of-the-art. Also, the ÜGK/COFO data analysts showed a convincing mastery of the psychometrics toolbox.

4. Standard-Setting

Thanks to the comprehensive and transparent documentation, the auditors were able to easily reconstruct and scrutinise the ÜGK/COFO standard-setting process. The applied bookmark standard-setting method is scientifically well-established and is a solid choice. The number of participants involved in the process was sufficient, and the group composition was well reflected: All language groups were equally represented, the group included people who were and people who were not involved in the item/test development process, and the group included didactics experts as well as practitioners. It is important to mention that the psychometricians acted only as moderators and/or observers, which again was a wise decision because they had no accredited expertise in mathematics content knowledge. Standard-settings are complex and complicated processes in which competent, structured, and yet diplomatic moderation is key. Given that it was the first time that such a standard-setting process took place in the context of the ÜGK/COFO, the decision to invite acknowledged external experts to observe the process—and if necessary to intervene—was the right call from a quality assurance perspective. The standard-setting process was well prepared, and the participants were adequately informed as well as trained. To the best of the auditors' knowledge, the bookmark method execution was exemplary at every stage in the process. The final decision to go for a single global mathematics scale was the right call given the proposed subscale cut-scores and the distribution of the item difficulties for the various subscales. The standard-setting process proved to be socially challenging, but this was to be expected given the high stakes of the operation and also given the "natural" underlying tensions between theorists and practitioners and the "traditional" tensions between language regions. The process also reportedly had a bridge-building effect at the very end. The idea to have the participants classify PISA mathematics items along the HarmoS minimum standards (Grundkompetenzen; compétences fondamentales)—a step that had not initially been foreseen in the process—was rather ingenious and proved to be very helpful to the auditors' understanding and interpretation of the 2016 ÜGK/COFO results.

To sum up, when all factors were taken into account, the 2016 ÜGK/COFO mathematics standard-setting can be considered scientific best practice and a textbook example of the bookmark standard-setting method.

5. Measurement of Minimum Standards

The ÜGK/COFO 2016 mathematics items were presented to the auditors in the form of a comprehensive and comprehensible item book. For all ÜGK/COFO test items, without exception, the auditors could comprehend and reconstruct their classification across the mathematics domains as well as with respect to their aspects as defined in the minimum standards (Grundkompetenzen; compétences fondamentales). In several cases, alternative classifications would have been acceptable. However, this nonexclusive classification is not an ÜGK/COFO specificity but a general condition of mathematics assessments, and thus, it is no reason for concern or criticism. Regarding the question of whether or not the 2016 ÜGK/COFO mathematics items operationalise the minimum standards with the intended level of difficulty, the auditors have to pass. A definite statement regarding the adequacy of item difficulty would be scientifically presumptuous because it would render the standard-setting processes ipso facto obsolete. This said, the auditors are comfortable in stating that they are under the strong impression that the items that were retained during the standard-setting procedure (i.e., the items below the cut-score) can be used to operationalise the theoretical descriptors in about the easiest way possible. In the same breath, the auditors would like to emphasise that many theoretical descriptors allow for the construction of test items of very different—not to say diametrically opposed—mathematical demands and difficulties, as the empirical difficulties of the ÜGK/COFO items unmistakably testify. Note that a proper empirical validation of the standards and/or an illustration of the latter can help to harmonise and streamline the interpretation and operationalisation of the theoretical descriptors. The auditors' conclusion regarding the descriptors' room for interpretation is, again, not a Swiss specificity but a more general issue in the definition and evaluation of mathematics competence.

What appears to be a genuinely Swiss specificity, however, are the partially very ambitious minimum standards (Grundkompetenzen; compétences fondamentales). While some theoretical descriptors potentially allow for the construction of relatively easy test items, other descriptors clearly do not. The ÜGK/COFO data aside—with barely two thirds of the students passing, the HarmoS mathematics minimum standards empirically behave more like norm standards (Regelstandards) rather than minimum standards (see definitions in the so-called "Klieme-expertise")—, the postulated comparably high demands of the Swiss minimum standards became particularly evident in the ÜGK standard-setting process. As already briefly mentioned, in a "bonus round", the standard-setting participants classified the 2012 PISA mathematics items according to the HarmoS minimum standards. The result of this classification is very instructive: almost all PISA Level 1, 2, and 3 items, about half of all the Level 4 and 5 items, and even a couple of Level 6 items were retained. Knowing that in the PISA assessment framework, Level 2 is considered to be the baseline level of proficiency needed to participate effectively and productively in a modern society and in future learning (i.e., the PISA minimum standard), one can only come to the conclusion that Swiss mathematics ambitions are very if not extremely high in an international comparison. It goes without saying that the comparison between PISA and ÜGK/COFO is far from ideal because the two assessments are conceptually very different. However, given the unambiguousness of the classification outcome, the auditors are persuaded that this exercise nevertheless perfectly illustrates the root of the "unexpected" 2016 ÜGK/COFO mathematics results. In ÜGK/COFO-related didactics literature, it is argued that Switzerland is not an average modern society, but it is rather a professionally as well as politically very demanding one. This line of argumentation is then used to justify the high mathematics expectations. The auditors agree in principle that there are good reasons for why the Swiss education requirements—and thus the minimum standards—may very well be at the upper end of an international comparison. At the same time, the auditors would like to remind readers that Switzerland is not the only highly developed economy participating in PISA.

In conclusion, with respect to content, the ÜGK/COFO 2016 mathematics items represent the minimum standards (Grundkompetenzen; compétences fondamentales) as currently defined in the reference documentation. However, minimum standards can be high, very high, or too high.

III. Summary and Recommendations

1. Summary

The 2016 ÜGK/COFO mathematics assessment must be seen as a complex, complicated, and high-risk pioneering endeavour in the landscape of Swiss education. After thorough and critical analysis, the auditors conclude that, thanks to 1.5 decades of large-scale assessment organisation and administration expertise built up through international large-scale studies (e.g., PISA), thanks to a traditionally very strong Swiss didactics expertise, thanks to a small but excellent cluster of psychometrics expertise, and, last but not least, thanks to substantial commitment and dedication of the many parties involved, the first ever ÜGK/COFO can—and should—be considered *summa summarum a success*. The auditors are convinced that the ÜGK/COFO data represent a long-term asset for Swiss educational decision-makers and researchers alike. The auditors are further confident that the ÜGK/COFO, in combination with TREE and FORS (but note that this list is by no means exhaustive), bears the potential to establish an internationally competitive Swiss cluster of excellence in large-scale assessments for the middle and long terms.

Regarding the three questions that guided the present investigation and evaluation (see *Context and Contractual Mission*), the auditors conclude, in a nutshell, that:

(1) Item development can be substantially improved and is currently the weakest link in the ÜGK/COFO operation. Although the item development processes are far from state-of-the-art, the outcome was nevertheless acceptable, the resulting assessment data were plausible, and the data were thus definitely interpretable (for more details, see *Item and Test Development*).

(2) Overall, the test design, data collection, data analysis, and, in particular the standard-setting are on a very high level and fully in line with present scientific quality standards (for more details, see *Test Administration*, *Data Processing and Analysis*, and *Standard-Setting*).

(3) With respect to content, the ÜGK/COFO 2016 mathematics items represent the minimum standards (Grundkompetenzen; compétences fondamentales) as currently defined in the reference documentation. The items retained during the standard-setting procedure (i.e., those items below the cut-score) operationalise the theoretical descriptors in about the easiest way possible. Crucially, however, these descriptors, and thus the minimum standards, are very ambitious in international comparison, lack proper empirical validation, and empirically behave more like norm standards (Regelstandards; for more details, see *Measurement of Minimum Standards*).

For easy reference, *Figure 1* summarises and organises the principal audit results in a comprehensible SWOT matrix.

Strengths	***Weaknesses***
test design data collection data analysis standard-setting documentation (except item dev.) didactics expertise political governance	item development translation processes assessment expertise involvement of practitioners empirical validation of standards cut-off for region DIF
acceptance by practitioners HarmoS framework assessment data TREE and FORS international LSA studies LSA competence centre	reference documents item development resources scientific governance project management assessment technology commercial players
Opportunities	***Threats***

Figure 1. ÜGK/COFO SWOT analysis

2. Recommendations

In order to further improve what is already a high-level operation, the auditors take the liberty to propose four, partially overlapping, and naturally nonbinding recommendations for future ÜGK/COFO (mathematics) operations:

(1) Revalue and revise the ÜGK/COFO item development. Test items, especially the closed-format items (e.g., multiple-choice items), have the disadvantage of generous face validity. In other

words, at first glance, a methodologically state-of-the-art assessment item easily resembles an amateurish question that only mimics test item properties. Over the last decade, a number of commercial assessment players conquered the parts of the Swiss education landscape that were not yet claimed by official stakeholders. It is the auditors' understanding that these players—or at least some of them—flooded the market with tempting products of questionable scientific validity. Such products all too often favour quantity over quality, especially when it comes to test items. Although it may be a long shot, the auditors are wondering whether the aforementioned circumstances did not at least implicitly endorse and propagate the idea that test items are uncomplicated, inexpensive, and easy to build. It is the auditors' impression that item development has been critically underestimated—not to say neglected—across the entire ÜGK/COFO setup. The underestimation is critical in the sense that test quality is largely defined by item quality, data quality substantially depends on test quality, and, last but not least, conclusions stand and fall with the data. In the present case, this ultimately affects HarmoS. In other words, aside from the education standards, the test items are the foundation of the ÜGK/COFO endeavour, and if there is a problem with the foundation, everything on top of it will fall sooner or later (see *Figure 2*). The stipulated "neglect" of item development is visible on several different levels: (a) In comparison with the context questionnaire—hyperbolically speaking, mere split variables for the ÜGK/COFO assessment data—item development barely exists in the KOSTA HarmoS protocols. (b) In comparison with data analysis or IT, no dedicated human resources have been allocated to item development, and it is expected that item development will happen in addition to the developers' regular work tasks. (c) Cum grano salis, translation processes for ÜGK/COFO administration manuals are more formalised than for ÜGK/COFO test items. Crucially, it is precisely a state-of-the-art translation process of test items (see e.g., PISA) that is conditio sine qua non for avoiding biased conclusions regarding HarmoS. (d) Although item development is also a matter of didactics, it is by no means *exclusively* a matter of

didactics. In ÜGK item development, solid didactics expertise is well represented. By contrast, explicit assessment and item writing expertise—as a matter of fact an expertise area of its own—is missing. Practitioners too are barely involved in the process. (e) As the KOSTA HarmoS protocols document, item development coordination and leadership were challenging. However, although duly noted, no substantial adjustments were made in the course of action. Running an operation such as ÜGK/COFO always involves a tightrope walk between different disciplines (i.e., didactics, psychometrics, pedagogics, statistics, logistics, IT, politics, and public relations). In the eyes of the auditors, the single biggest challenge in an operation of the ÜGK/COFO type resides in finding the golden path—the common denominator that joins all areas without sacrificing any. In other words, ÜGK/COFO is by design and definition a collaborative game of compromise that requires bridge builders rather than disciplinary memorial architects.

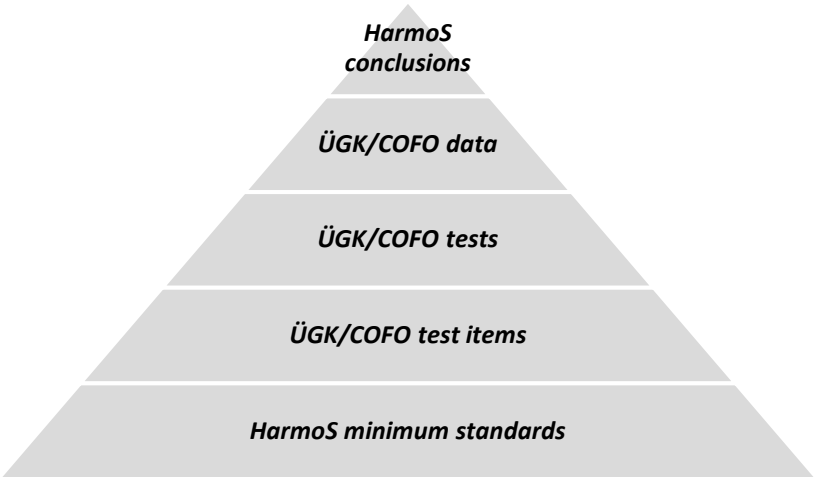

Figure 2. HarmoS/ÜGK/COFO hierarchy

Regarding the revaluation and revision of the ÜGK/COFO item development, the auditors concretely recommend to fundamentally rethink the resources allocated to item development. In comparison, roughly two thirds of all LUCET resources—human as well as financial—go into item and test development. As a matter of fact, item and test development is the single most expensive part of the Luxembourg school monitoring programme. The auditors further recommend to (a) link item development more closely to test development and data analysis, which are two clear ÜGK/COFO

assets, (b) strengthen the coordination mandate of the ADB/BDT cluster, (c) complement the item development working groups with practitioners as well as acknowledged assessment experts, (d) thoroughly pretest all materials, and, last but not least, (e) invest in rigorous item translation processes. The 2016 ÜGK/COFO language region DIF cut-off is at the very upper end of DIF generosity. Item developers could exploit the graphical possibilities of the new test medium and develop qualitatively better, less text-heavy test items; these will allow for a less lenient language region DIF cut-off and thus eventually for more valid and reliable conclusions regarding HarmoS. Also, while there is a clear need for investment in item development, there is also room for savings. Although the ÜGK/COFO test design is state-of-the-art, the number of items—and thus also the number of rotating clusters—could be reduced. The 2016 ÜGK/COFO item quantity was very generous for a test that, in the end, revolves around a single cut-score and is not used for individual diagnostics. The auditors further advocate for a general reduction of (half)open item formats because expertly designed closed-format items can also assess very complex processes with better standardisation and far better test economy in comparison with (half)open questions.

(2) Reconsider the minimum standards (Grundkompetenzen; compétences fondamentales) in mathematics. It is the conclusion and conviction of the auditors that the "unexpected" 2016 ÜGK/COFO mathematics results, which at least implicitly triggered the present audit, ultimately have their root in suboptimal and overly ambitious reference documents. Before outlining several possible strategies for how to approach this delicate matter, the auditors would like to recall the reason that initially led to the decision to choose minimum standards as the conceptual weapon of choice for educational harmonisation. In the auditors' understanding of the situation, minimum standards were primarily chosen to explicitly include low-achieving students. That said, by defining overly ambitious minimum standards, it is precisely these low-achieving students—as well as their teachers—who will be left out, as even the most basic targets will be (realistically speaking) out

of reach. At least three strategies can be applied to tackle the conundrum of mathematics minimum standards that are reached by barely two thirds of the population. These will be presented in the order of scientifically most to least appealing:

(a) The empirical approach. Use the 2016 ÜGK/COFO mathematics data to empirically validate and revise the mathematics minimum standards (Grundkompetenzen; compétences fondamentales). Although the HarmoS competence model underwent a validation study, the derived minimum standards were never properly empirically validated. On the basis of the HarmoS competence model, the auditors tried hard to understand how *exactly* the minimum standards emerged, but neither the available documentation nor the interviewees could provide a scientifically satisfying answer. Moreover, the aforementioned validation study was not without its methodological flaws and represents only the first step in an iterative validation process. These two facts were also explicitly acknowledged and emphasised by the HarmoS methodology group in their solid final report. The capital advantage of this first strategy is that it should ultimately produce didactically as well as empirically sound mathematics minimum standards—a prerequisite for a state-of-the-art ÜGK/COFO (see also *Figure 2*) as originally intended in the HarmoS white book. Moreover, in considering the empirical reality of the assessment, the practitioners' acceptance of the precious ÜGK/COFO will not be jeopardised. The disadvantages are that the reference documents—standards as well as curricula—will require adaptations and that the 2016 ÜGK/COFO results cannot be published as planned.

(b) The terminological approach. With respect to content, stick to the current standards and descriptors but revise the overarching terminology and refer henceforth to the mathematics standards as "norm standards" (Regelstandards) instead of "minimum standards". It is important to mention that this terminological change will not hinder the harmonisation or evaluation of the latter. However, it requires a change of perspective in the ÜGK/COFO: Instead of investigating what *every* Swiss student should be able to do, the assessment would henceforth determine what *the average*

Swiss student should be able to do. The advantages of this second approach are that mathematics standards and curricular documents can largely remain as they are, and the 2016 ÜGK/COFO results can be made public as planned. Aside from some potential challenges in communicating the results, the biggest disadvantage here will be incoherence between school subjects, because, to the best of the auditors' knowledge, the minimum standards for language, tested in the 2017 ÜGK/COFO, empirically behave as would be expected. Thus, why would norm standards be defined and measured for one subject when minimum standards are measured for the others?

(c) The rhetorical approach. Argue that the 2016 ÜGK/COFO mathematics results were as expected because the participants were not yet being explicitly taught on the basis of the HarmoS minimum standards (Grundkompetenzen; compétences fondamentales) and the associated curricula. The advantages of this third approach are identical to those of the second strategy. The disadvantages are that the proposed line of argumentation would stand on relatively shaky ground as sceptics would be likely to ask: Are the current mathematics curricula really *fundamentally* different from previous ones? Also, this solution would offer only a one-time explanation, and, consequently, the issue might just be postponed to the next ÜGK/COFO mathematics assessment.

Last but not least, and independent of the choice of strategy, in order to fully understand the picture painted by the 2016 ÜGK/COFO, it is of utmost importance to zoom out, widen the discussion, and rigorously analyse and debate the assessment data in due consideration of the ÜGK/COFO context questionnaire data.

(3) Re-evaluate the overall management and coordination of the ÜGK/COFO. The EDK/CDIP currently wears two mutually exclusive hats: as the contracting and supervising authority on the one hand, and as the ÜGK/COFO project coordination team—and thus executive body—on the other hand. Crucially, while the contracting and supervising authority is by definition a political body, the executive body should be apolitical and scientific in nature. Although the two bodies ideally form a symbiotic relationship, occasional tensions and even conflicts are naturally to be expected.

In order to find the best possible compromise in the eventuality of a conflict—and to make the best possible project decisions in general—it is of utmost importance that the two bodies are independent, adequately represented, and engaged in regular, formalised, and transparent dialogue. The auditors fully understand that the current organisational setup was the logical choice for the ÜGK/COFO incubation phase. However, at this point in time, the project has outgrown its setup phase, and a clear separation of roles will need to be imposed if past efforts are to be consolidated, if the ÜGK/COFO is going to evolve, and if the ÜGK/COFO is going to be coordinated and supervised at the highest professional level.

(4) Complement the well-functioning strategic/political ÜGK/COFO governance with a scientific sister-body. The ÜGK/COFO currently lacks a scientific authority that overlooks all aspects of the project. In order to ensure the scientific state-of-the-art in the long term and to consult supervising and executive bodies alike, ÜGK/COFO urgently requires the implementation of the long foreseen scientific consortium (WiKo). Given the fact that large-scale assessment is still in pioneering territory in the Swiss education panorama, the auditors strongly suggest to complement this scientific consortium—in the auditors' reading of the ÜGK/COFO documentation a purely national body that may also be directly involved in certain ÜGK/COFO processes—with a scientific advisory board of renowned (inter)national experts that are *fully* independent of the ÜGK/COFO operation, and have no own stakes in the latter. It would be important for the expertise represented in the scientific consortium as well as in the proposed scientific advisory board to cover all aspects of the ÜGK/COFO operation.

Index of Acronyms and Abbreviations

ADB	*Aufgabendatenbank*
ADMEE	*Association pour le Développement des Méthodologies d'Évaluation en Éducation*
BDT	*Banque de données de tâches*
BELDACH	*Belgien-Luxemburg-Deutschland-Österreich-Schweiz*
CCI	*Chambre de commerce et d'industrie*
CDIP	*Conférence suisse des directeurs cantonaux de l'instruction publique*
CH-D	*Deutschschweiz*
CH-F	*Suisse romande*
CIIP	*Conférence intercantonale de l'instruction publique*
COFO	*Vérification de l'atteinte des compétences fondamentales*
CSRE	*Centre suisse de coordination pour la recherche en éducation*
DACHL	*Deutschland-Österreich-Schweiz-Luxemburg*
DIF	*Differential item functioning*
EDK	*Schweizerische Konferenz der kantonalen Erziehungsdirektoren*
ÉpStan	*Épreuves Standardisées*
ETS	*Educational Testing Service*
FORS	*Fondation suisse pour la recherche en sciences sociales*
GS	*Generalsekretariat*
HarmoS	*Harmonisierung der obligatorischen Schule*
HBSC	*Health Behaviour in School-aged Children*
ICC	*Item characteristic curve*
ICCS	*International Civic and Citizenship Education Study*
IRT	*Item response theory*
IT	*Information technology*
KOSTA	*Koordinationsstab*
LSA	*Large-scale assessment*
LUCET	*Luxembourg Centre for Educational Testing*
OASYS	*Online Assessment System*

PDF	*Portable Document Format*
PH FHNW	*Pädagogischen Hochschule Fachhochschule Nordwestschweiz*
PH SG	*Pädagogische Hochschule St. Gallen*
PIRLS	*Progress in International Reading Literacy Study*
PISA	*Programme for International Student Assessment*
SG	*Secrétariat général*
SKBF	*Schweizerische Koordinationsstelle für Bildungsforschung*
SWOT	*Strengths-Weaknesses-Opportunities-Threats*
TAO	*Testing Assisté par Ordinateur*
TREE	*Transitionen von der Erstausbildung ins Erwerbsleben*
ÜGK	*Überprüfung des Erreichens der Grundkompetenzen*
Uni BE	*Universität Bern*
Uni GE	*Université de Genève*
WiKo	*Wissenschaftliches Konsortium*
WLE	*Weighted likelihood estimation*

Appendix A – Stakeholder Hearings

Angelone, D. (with Keller, F.)	*SKBF/CSRE – ÜGK/COFO ADB/BDT, test dev., data analysis & standard-setting co-lead*	25.01.	Bern
Dorier, J.-L.	*Uni GE – item dev. CH-F lead*	23.01.	Skype
Erzinger, A.	*PH SG – ÜGK/COFO administration CH-D lead*	25.01.	Bern
Füeg, J. (with Husfeldt, V.)	*GS EDK/SG CDIP – ÜGK/COFO project lead (support)*	25.01.	Bern
Hascher, T.	*Uni BE – Setup ÜGK/COFO scientific consortium (WiKo) lead*	29.01.	Skype
Husfeldt, V. (with Füeg, J.)	*GS EDK/SG CDIP – ÜGK/COFO project lead*	25.01.	Bern
Keller, F. (with Angelone, D.)	*SKBF/CSRE – ÜGK/COFO ADB/BDT, test dev., data analysis & standard-setting co-lead*	25.01.	Bern
Linneweber-Lammerskitten, H.	*PH FHNW – ÜGK/COFO item dev. lead & item dev. CH-D lead*	25.01.	Bern
Maradan, O.	*SG CIIP – formerly dev. of HarmoS minimum standards lead*	25.01.	Bern
Scherrer, M.	*Teacher CH-D – standard-setting participant*	25.01.	Bern
Schönenberger, S.	*PH SG – ÜGK/COFO coding of (half)open test items lead & standard-setting participant*	26.01.	Skype

Appendix B – Documentation

Aktennotiz zur Sitzung 1/16 der Projektleitung ÜGK *[20160108_pl-ugk_protokoll.pdf]*

Aktennotiz zur Sitzung 2/16 der Projektleitung ÜGK *[20160201_pl-ugk_protokoll.pdf]*

Aktennotiz zur Sitzung 3/16 der Projektleitung ÜGK *[20160301_pl-ugk_protokoll.pdf]*

Aktennotiz zur Sitzung 4/16 der Projektleitung ÜGK *[20160401_pl-ugk_protokoll.pdf]*

Allgemeiner Rahmen für die Schlussüberarbeitung der nationalen Standards *[100910Vorgaben-Korr-Standards.pdf]*

Am Projekt HarmoS beteiligte Personen und Institutionen Auszug aus dem internen Schlussbericht *[080115InternerSchlussbericht72Am_Projekt_HarmoS_beteiligte_Personen_und_Institutionen.pdf]*

Arbeitsablauf zur Itemerstellung *[140522_Arbeitsablauf_Itemerstellung_bgi.pdf]*

Basisstandards für die Mathematik: Konsultationsdossier *[090820Basisstandards-Mathe-d-20-08-09.pdf]*

Beschluss Plenarversammlung 12. Juni 2014: Einsatz von externen Testleitenden / Kodierenden im Rahmen der Überprüfung des Erreichens der Grundkompetenzen: Beschlussfassung *[PB_grundkompetenzen_testleitende_d.pdf]*

Beschluss Plenarversammlung 20. Juni 2013: Überprüfung der Erreichung der Grundkompetenzen; Konzept: Verabschiedung *[PB_grundkompetenzen_d.pdf]*

Beschluss Plenarversammlung 22. Juni 2017: Überprüfung des Erreichens der Grundkompetenzen: Planung zu weiteren ÜGK-Erhebungen ab 2020 *[pb_grundkompetenzen_erhebungen_2020_d.pdf]*

Beschluss Plenarversammlung 23. Juni 2016: Fortführen der Aufgabendatenbank EDK ab 2017: Verabschiedung *[pb_Adb_d.pdf]*

Beschluss Plenarversammlung 25. Oktober 2012: Überprüfung der Erreichung der Grundkompetenzen; Konzept und Finanzierung der Aufgabendatenbank: Beschlussfassung *[PB_aufgabendatenbank_d.pdf]*

Bildungsstandards in der Schweiz. Grundkompetenzen HarmoS Mathematik Jahrgangsstufe 11 (2 Jahre Vorschule & 9 Jahre Schule) *[141101HarmoSMatrixJh11Dt.pdf]*

Contrat entre la CDIP et la PH FHNW concernant le développement d'items et de tests de mathématique *[2014_vereinbarung_aufgabenentwicklung_mathematik_p.pdf]*

Dokumentation ÜGK – Plenarversammlung, Vorstand, Kosta HarmoS, Verträge *[GK_Dokumentation_Gremien.pdf]*

Empfehlung des Gremiums zur Schwellenwertsetzung an den KOSTA HarmoS *[2_20170630_kosta_harmos_empfehlung_gremium_schwell.pdf]*

Entwurf zur Organisation und Konzeption der Aufgabendatenbank HarmoS Mathematik *[140528V05OrganisationUndKonzeptionDatenbankHarmoSMathematik.pptx]*

Ergebnisse aus dem Anhörungsprozess / der Konsultation und weitere Arbeiten bis zur Verabschiedung der Standards *[100913AnhoerungKonsultation.pdf]*

Fachdidaktische Einschätzung (in Form eines Kurzberichts) zur Vergleichbarkeit der Konstrukte mathematischer Kompetenzen in PISA 2012/15 und ÜGK 2016 sowie zur Vergleichbarkeit der Testaufgaben in PISA 2012/15 und ÜGK 2016 hinsichtlich ihrer formalen und sprachlichen Anforderungen *[171009V35Vergleichbarkeit_PISA_UEGK.pdf]*

Faktenblatt Nationale Bildungsziele für die obligatorische Schule: in vier Fächern zu erreichende Grundkompetenzen *[grundkomp_faktenblatt_d.pdf]*

Fragestellungen an die Teilnehmenden der Vernehmlassung über die HarmoS-Basisstandards *[080124Fragen_Vernehmlassung_Standards-de.pdf]*

Geschäftsstelle Aufgabendatenbank: Aufgabenentwicklung Schul-/Fremdsprachen und Mathematik *[Kosta_160407_1.pptx]*

Grundinformationen zu den HarmoS-Bildungsstandards: Kurze Einführung zur Vernehmlassung 2008 über die Vorschläge für Basisstandards *[080110EINFÜHRUNG-Vernehml-Standards.pdf]*

Grundkompetenzen für die Mathematik *[grundkomp_math_d.pdf]*

Grundkompetenzen Obligatorische Schule: Informationen zur Durchführung der Erhebung in Sprachen (Ende Primarstufe) *[2016_10_infoblatt_grundkompetenzen_d.pdf]*

Grundkompetenzen Obligatorische Schule: Informationen zur Durchführung der Grundkompetenzen -Tests in Mathematik (Ende obligatorische Schule) und in Sprachen (Ende Primarstufe) *[151816_1_Infoblatt_d_p1-4.pdf]*

HARMOS Zielsetzungen und Konzeption Juni 2004 *[weissbuch_d.pdf]*

HarmoS: Zum methodologischen Vorgehen *[HarmoS_Weg_dt.pdf]*

Haupterhebung 2016 Mathematik: Anleitung für den Web-Readiness-Test *[ÜGK_M16_Anleitung_Web-Readiness.pdf]*

Haupterhebung 2016 Mathematik: Anleitung für die schulverantwortliche Person (SV) *[ÜGK_M16_AnleitungSV.pdf]*

Haupterhebung 2016 Mathematik: Anleitung für die testverantwortliche Person (TV) *[ÜGK_M16_AnleitungTV.pdf]*

Haupterhebung 2016 Mathematik: Skript für die testverantwortliche Person (TV) *[ÜGK_M16_SkriptTV.pdf]*

Haupterhebung 2016 Mathematik: Technische Anleitung *[ÜGK_M16_Technische Anleitung.pdf]*

Instrumente für die Überprüfung der Erreichung der Grundkompetenzen *[140217 WS I-3 Linneweber Instrumente Grundkompetenzen.ppt]*

Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS-Konkordat); Umsetzung auf der Ebene der interkantonalen Koordination vom 25./26. Oktober 2007 *[Beschluss_d.pdf]*

Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS-Konkordat) vom 14. Juni 2007 *[HarmoS_d.pdf]*

Konkordat über die Schulkoordination vom 29. Oktober 1970 *[1-1d.pdf]*

KOSTA HarmoS Ausschuss: Protokoll der Sitzung 10 vom 7. September 2016, 10.15 – 15.45 Uhr *[20161118_Kosta_Ausschuss_02_Protokoll-Nr-10.pdf]*

KOSTA HarmoS Ausschuss: Protokoll der Sitzung 2 vom 8. Januar 2015, 10.15 – 13.15 Uhr *[1_Protokoll_Kosta_Ausschuss_20150108.pdf]*

KOSTA HarmoS Ausschuss: Protokoll der Sitzung 3 vom 6. März 2015, 10.15 – 15.00 Uhr *[20150306_Protokoll_UEGK_Kosta_Ausschuss.pdf]*

KOSTA HarmoS Ausschuss: Protokoll der Sitzung 4 vom 19. Juni 2015, 10.15 – 12:30 Uhr *[20150619_Protokoll_Kosta_Ausschuss.pdf]*

KOSTA HarmoS Ausschuss: Protokoll der Sitzung 5 vom 20. August 2015, 13.15 – 16.45 Uhr *[20150820_Protokoll_Kosta_Ausschuss.pdf]*

KOSTA HarmoS Ausschuss: Protokoll der Sitzung 6 vom 6. November 2015, 10.15 – 16.45 Uhr *[20151106_Protokoll_Kosta_Ausschuss.pdf]*

KOSTA HarmoS Ausschuss: Protokoll der Sitzung 7 vom 20. Januar 2016, 10.15 – 15.00 Uhr *[20160120_Kosta_Ausschuss_AN.pdf]*

KOSTA HarmoS Ausschuss: Protokoll der Sitzung 8 vom 9. März 2016, 10.15 – 15.00 Uhr *[20160309_kosta_ausschuss_00_protokoll.pdf]*

KOSTA HarmoS Ausschuss: Protokoll der Sitzung 9 vom 8. Juni 2016, 10.15 – 15.15 Uhr *[20160608_kosta_ausschuss_00_protokoll.pdf]*

KOSTA HarmoS Ausschuss: Protokoll der Sitzung 9 vom 8. Juni 2016, 10.15 – 15.15 Uhr *[20160907_Kosta_Ausschuss_02_Protokoll-Nr-9.pdf]*

KOSTA Harmos: Aktennotiz Analyse der Belastbarkeit der ÜGK-Ergebnisse: Systematisierung der Fragestellungen *[02_20171124_an_kosta_uegk_systematisierung_fragest-3.pdf]*

KOSTA Harmos: Aktennotiz Aufgabendatenbank der EDK: Festlegung der Schwellenwerte *[3_20160915_an_kostaharmos_adb_schwellenwertsetzung.pdf]*

KOSTA Harmos: Aktennotiz Darstellung der ÜGK-Resultate *[6_20161108_tischpapier_angepasste_an_darstellung_u-3.docx]*

KOSTA Harmos: Aktennotiz Überprüfung des Erreichens der Grundkompetenzen: Kurzbericht zu den Erhebungen 2016 und 2017 *[02_20170915_kosta_harmos_bericht_uegk_2016_2017.pdf]*

KOSTA Harmos: Aktennotiz Überprüfung des Erreichens der Grundkompetenzen: Validierungsprozess der Kompetenzmodelle *[03_20170915_an_kosta_harmos_validierung.pdf]*

KOSTA HarmoS: Protokoll der Sitzung 32 vom 8. Januar 2015, 13.15 – 17.00 Uhr *[01_Protokoll_Kosta_20150108.pdf]*

KOSTA HarmoS: Protokoll der Sitzung 33 vom 2. April 2015, 10.15 – 15.45 Uhr *[01_20150402_Protokoll_Kosta.pdf]*

KOSTA HarmoS: Protokoll der Sitzung 34 vom 24. Juni 2015, 10.15 – 15.45 Uhr *[01_Protokoll_Kosta_20150624.pdf]*

KOSTA HarmoS: Protokoll der Sitzung 35 vom 4. September 2015, 10.15 – 15.30 Uhr *[1_Protokoll_Kosta_20150904.pdf]*

KOSTA HarmoS: Protokoll der Sitzung 36 vom 27. November 2015, 10.15 – 15.45 Uhr *[protokoll_kosta_20151127_v2.pdf]*

KOSTA HarmoS: Protokoll der Sitzung 37 vom 21. Januar 2016, 10.15 – 15.50 Uhr *[01_20160121_protokoll_kosta_harmos.pdf]*

KOSTA HarmoS: Protokoll der Sitzung 38 vom 7. April 2016, 10.15 – 15.45 Uhr *[1_Protokoll_Kosta_20160407.pdf]*

KOSTA HarmoS: Protokoll der Sitzung 39 vom 24. Juni 2016, 10.15 – 12.30 Uhr *[1_20160624_Kosta_Harmos_Protokoll.pdf]*

KOSTA HarmoS: Protokoll der Sitzung 40 vom 15. September 2016, 10.15 – 15.15 Uhr

[20160915_kosta_harmos_protokoll.pdf]

KOSTA HarmoS: Protokoll der Sitzung 41 vom 25. November 2016, 10.15 – 13.00 Uhr *[01_protokoll_kosta-harmos_20161125.pdf]*

Kurzbericht zur Entwicklung mathematischer Testitems für HarmoS / ÜGK 2016 *[180125V14Kurzbericht_Itementwicklung_HarmoS_UEGK.docx]*

Kurzdokumentation zur ÜGK Aufgaben-entwicklung und Testzusammenstellung *[Item- und Testentwicklung_ÜGK2016_031215.pdf]*

Kurz-Info Obligatorische Schule: Schulstufen, Zählweise der Schuljahre *[kurzinfo_zaehlweise_d.pdf]*

Kurzprotokoll der Sitzung 10/16 der Projektleitung ÜGK *[20161026_pl-ugk_protokoll.pdf]*

Kurzprotokoll der Sitzung 11/16: der Projektleitung ÜGK *[20161108_pl-ugk_protokoll.pdf]*

Kurzprotokoll der Sitzung 12/16: der Projektleitung ÜGK *[20161207_pl-ugk_protokoll_.pdf]*

Kurzprotokoll der Sitzung 5/16 der Projektleitung ÜGK *[20160428_pl-ugk_protokoll.pdf]*

Kurzprotokoll der Sitzung 6/16 der Projektleitung ÜGK *[20160523_pl-ugk_protokoll.pdf]*

Kurzprotokoll der Sitzung 7/16 der Projektleitung ÜGK *[20160620_pl-ugk_protokoll.pdf]*

Kurzprotokoll der Sitzung 8/16 der Projektleitung ÜGK *[20160816_pl-ugk_protokoll.pdf]*

Kurzprotokoll der Sitzung 9/16 der Projektleitung ÜGK *[20160913_pl-ugk_protokoll.pdf]*

Large-Scale Assessments in der Schweiz: Prozesshandbuch zur Durchführung von Erhebungen *[20171121_prozesshandbuch_lsa_schweiz_v18.pdf]*

Lehrplan 21 Gesamtausgabe *[V_EF_DE_Gesamtausgabe.pdf]*

Leistungsvereinbarung zwischen EDK und PH FHNW betreffend die Beratung bei der technischen Weiterentwicklung der Aufgabendatenbank für die Aufgaben- und Testentwicklung im Bereich Mathematik *[vereinbarung_phfhnw_techn _weiterentwicklung_def.pdf]*

Liste aller Dossiers, die zum Review verschickt wurden Stand: 30.03.15 11.30Uhr *[150330ListeAllerDossiersReview[1]*.pdf]

Mathematik Projekt HarmoS M_d_9_001 *[070307TestheftM_d_9_001.pdf]*

Mathematikdidaktik, Bildungsstandards und mathematische Kompetenz (in Fachdidaktik Mathematik: Grundbildung und Kompetenzaufbau im Unterricht der Sek. I und II) *[Scan Fachdidaktik Mathematik Kap 1.pdf]*

Medienmitteilung Die Ziele der obligatorischen Schule harmonisieren: Anhörungsprozess zu nationalen Bildungszielen *[100125BildungszieleMedienmitteilung_d.pdf]*

Organisation und Verlauf der Korrekturen *[Seiten aus 080115Schlussbericht72-3.pdf]*

Organisationsreglement für die Aufgabendatenbank der Schweizerischen Konferenz der kantonalen Erziehungsdirektoren vom 23. Januar 2014 *[fj_Organisationsreglement _d_140123.pdf]*

Organisationsreglement über die Durchführung der Überprüfung des Erreichens der Grundkompetenzen vom 8. Mai 2014 *[Organisationsregl_UeGK_d.pdf]*

Piloterhebung Mathematik 2015 Email *[150917EmailAngelone.pdf]*

Plan d'études romand *[PER_complet.pdf]*

Projekt-Organigramm der Überprüfung des Erreichens der Grundkompetenzen (ÜGK) *[02_20160209_organigramm_ugk_df.pdf]*

Reglement des Koordinationsstabes für die Umsetzung der Interkantonalen Vereinbarung über die Harmonisierung der obligatorischen Schule (Kosta HarmoS) vom 24. Januar 2008 *[KostaHarmos_d.pdf]*

Schlussbericht der HarmoS-Methodologiegruppe *[03_methodologie-schlussbericht-_harmos11.pdf]*

Systemerweiterung für die Kodierung ÜGK Schlussbericht *[20170410_Codiersystem_UeGK_V1.pdf]*

Test principal 2016 pour les mathématiques : Manuel des Coordinateurs scolaires (CS) *[COFO_M16_Manuel-CS_fra.doc]*

Test principal 2016 pour les mathématiques : Manuel pour le test de compatibilité web *[COFO_M16_Manuel_Test-compatibilite-web_fra.pdf]*

Test principal 2016 pour les mathématiques : Manuel pour les administrateurs de test (AT) *[COFO_M16_Manuel-AT_fra.pdf]*

Test principal 2016 pour les mathématiques : Manuel technique pour les administrateurs de test *[COFO_M16_Manuel-technique_fra.docx]*

Test principal 2016 pour les mathématiques : Script pour les administrateurs de test (AT) *[COFO_M16_Script_AT_fra.docx]*

Überprüfung der Erreichung der Grundkompetenzen in Mathematik: Zwischenbericht zum Aufbau der Aufgabendatenbank *[150108Zwischenbericht.pptx.pdf]*

ÜGK 2016 Mathematik HarmoS 11: Technische Dokumentation zur Schwellenwertsetzung *[Schwellenwertsetzung_ÜGK2016_Entwurf_1.pdf]*

ÜGK 2016 Mathematik HarmoS 11: Technische Dokumentation zur Testentwicklung und Skalierung *[Testdesign_Skalierung_ÜGK2016__Entwurf_1.1.pdf]*

ÜGK 2016 Mathematik, 11. Klassenstufe: Standard-Setting *[3_20160915_kosta_ausschuss_konzept_standardsetting.pdf]*

ÜGK 2016 Mathematik, 11. Klassenstufe: Standard-Setting *[20160816_pl-ugk_beilage_adb_standardsetting.pdf]*

ÜGK Mathematik 2016, HarmoS 11: Itembook *[Itembook_Mathematik_ÜGK2016_final.pdf]*

ÜGK Piloterhebung Mathematik 2015: Bericht zur Itemanalyse *[Itemanalyse_Pilotherhebung_Mathematik_ÜGK2016.pdf]*

ÜGK Piloterhebung Mathematik 2015: Technischer Bericht zur Itemanalyse *[GK_FT_2015_Erste Ergebnisse Itemnalayse_Version1.pdf]*

ÜGK Pilottest 2016: Ergebnisse der Itemanalyse *[20161026_pl-ugk_beilage4_prasentation_itemanalyse.pdf]*