

Klassencockpit

Evaluation

im Auftrag der Projektleitung Klassencockpit

Zürich, 27. Oktober 2002

Inhalt

1. Gegenstand der Evaluation und Auftrag	3
2. Durchführung	4
2.1 Populationen und Stichprobenziehung.....	4
2.2 Art und Einsatz der Befragungsinstrumente.....	4
2.3 Teilnahmequoten	5
2.4 Datenauswertung	5
2.5 Aussagekraft der Ergebnisse.....	5
3. Ergebnisse	7
3.1 Bekanntheit und Verbreitung des Klassenscockpit (Repräsentativbefragung)	7
3.2 Anwendung des Klassenscockpit	11
3.3 Qualitative Ergebnisse	14
4. Zusammenfassung, Diskussion und Folgerungen	21
5. Empfehlungen	25

1. Gegenstand der Evaluation und Auftrag

Klassencockpit ist eine standardisierte Aufgabensammlung in Deutsch und Mathematik für 3. bis 9. Klassen der Volksschule. Pro Jahr können bis drei Module zu einem vorgegebenen Zeitpunkt eingesetzt werden. Die Schülerdaten können online eingegeben werden; die Auswertung erfolgt für die einzelnen SchülerInnen, die Klasse und im Vergleich zu einer kantonalen Vergleichsstichprobe von 25-30 zufällig ausgewählten Klassen. Klassencockpit beruht auf der Idee, dass aufgrund des Vergleichs mit anderen Klassen eine Standortbestimmung vorgenommen und anhand der analysierten Leistungen der Unterricht gezielt geplant werden kann.

Mit dem Ziel, den bisherigen Einsatz kritisch zu hinterfragen und dadurch Hinweise auf die Stärken und Verbesserungsnotwendigkeiten des Instrumentes zu erhalten, wurde eine externe Evaluation in Auftrag gegeben. Im Einzelnen sollten folgende Aspekte behandelt werden:

- Verbreitung und Einsatzhäufigkeit von Klassencockpit
- Akzeptanz des Instrument in der Lehrerschaft und Einsatzzwecke
- Allgemeine Beurteilung der Nützlichkeit und Eruiere von Problembereichen

Anhand von Befragungsdaten sollten Empfehlungen formuliert bzw. Hinweise auf Bereiche zur Weiterentwicklung/Modifikation von Klassencockpit gegeben werden.

Der Auftrag wurde bearbeitet von Charles Landert (Konzept, Befragungsinstrumente, Interviews, Berichterstattung), Felix Davatz (statistische Analyse) und Barbara Landert (Datenerfassung, Codierung offene Fragen).

2. Durchführung

2.1 Populationen und Stichprobenziehung

Die Evaluation bediente sich zweier Datenquellen:

- Population 1: Mittel- und Oberstufenlehrpersonen im Kanton St. Gallen
- Population 2: Klass Cockpit-Nutzerinnen und -Nutzer im Kanton St. Gallen

Aus diesen beiden, teilweise sich überschneidenden Grundgesamtheiten wurde je eine Stichprobe gezogen:

- *St. Galler Mittel- und Oberstufenlehrpersonen:* Das Erziehungsdepartement des Kantons SG stellte einen aktuellen Adressdatensatz zur Verfügung. Unter den insgesamt 2'130 Adressen waren 955 (44.9%) von Mittelstufenlehrpersonen und 1'175 (55.1%) von Oberstufenlehrpersonen. Unter Einhaltung dieser Proportionen wurden 370 Mittelstufenlehrpersonen beziehungsweise 454 Oberstufenlehrpersonen im Zufallsverfahren der Stichprobe ($N_{\text{total}} = 824$) zugeteilt und mit dem schriftlichen Fragebogen bedient.
- *Nutzerinnen und Nutzer des Klass Cockpit:* Aus dem Datensatz des Kompetenzzentrums für Bildungsevaluation und Leistungsmessung an der Universität Zürich (KBL) wurden 80 Adressen von Nutzerinnen und Nutzern des Klass Cockpits zufällig ausgewählt. Nach dem Abgleich mit der Stichprobe aller Mittel- und Oberstufenlehrpersonen standen für die qualitative Befragung Adressen von 27 Mittelstufen- und 29 Oberstufenlehrpersonen zur Verfügung, von denen ein Teil telefonisch kontaktiert und anhand eines Gesprächsleitfadens qualitativ befragt wurde.

2.2 Art und Einsatz der Befragungsinstrumente

Für die *schriftliche Befragung* der Repräsentativstichprobe wurde ein vierseitiger Fragebogen mit insgesamt 15 Fragen mit strukturierten Antwortvorgaben und 2 offenen Fragen abgegeben.

Den *qualitativen Gesprächen* mit den Nutzerinnen und Nutzern lag ein Gesprächsleitfaden zugrunde, der die relevanten Befragungsthemen in Stichworten festhielt. Der Leitfaden bildete den Ausgangspunkt und das Gerüst der Befragung. Situativ wurde er ergänzt mit Anschlussfragen, die entweder der Präzisierung, Vertiefung oder Verifizierung abgegebener Antworten dienten. Da vor dem Start der qualitativen Befragung die schriftliche Befragung bereits provisorisch ausgewertet war, konnten die Gespräche auf Ergebnisse abgestützt erfolgen und sich fallweise auch auf einzelne Themenbereiche (Einsatzzwecke, Lehrplankompatibilität, Noten usw.) konzentrieren.

2.3 Teilnahmequoten

Die Bereitschaft zur Teilnahme war in beiden Teilevaluationen erfreulich hoch.

- Bei der *schriftlichen Befragung* wurden (von erwarteten 50-70%) 444 oder netto 55%¹ der versandten Fragebogen retourniert. Total 435 rechtzeitig eingetroffene Fragebogen konnten ausgewertet werden.
- Bei der *qualitativen telefonischen Nutzerbefragung* mussten insgesamt 42 Personen kontaktiert werden, um 38 Gespräche zu realisieren. Dies entspricht einer sehr guten Teilnahme von 90%. Wer ein Interview ablehnte, hatte dafür Gründe, die inhaltlich nichts mit dem Untersuchungsthema zu tun hatten (Überlastung, Unabkömmlichkeit, Urlaub).

Die unterschiedlichen Teilnahmequoten können weitgehend durch die Befragungsmethode (direkte Ansprache am Telefon vs. schriftliche Befragung mit quasi-anonymem Absender) erklärt werden.

2.4 Datenauswertung

Schriftliche Befragung:

Nach ihrem Eingang wurden die ausgefüllten Fragebogen kontrolliert und die auf offene Fragen gegebenen Antworten codiert. Im Anschluss an diese vorbereitenden Arbeiten wurden die Daten elektronisch erfasst, mit einschlägiger Statistiksoftware ausgewertet und tabellarisch dargestellt. Die inhaltliche Interpretation der Daten bildete den Abschluss dieses Auswertungsvorgangs.

Qualitative telefonische Befragung:

Die leitfadengestützten Gespräche wurden stichwortartig protokolliert und qualitativ ausgewertet. Da die meisten qualitativen Gespräche nach der schriftlichen Befragung realisiert wurden, wurden die Ergebnisse anhand des vorhandenen Auswertungsrasters eben dieser schriftlichen Befragung ausgewertet und im Auswertungsteil *Qualitative Ergebnisse* integriert. Die Absicht, die Stichprobe nach TeilnehmerInnen mit unterschiedlichem Nutzungsgrad zu differenzieren, wurde mit Blick auf die sehr unterschiedliche Häufigkeit entsprechender Teilgruppen und somit aus forschungsökonomischen Gründen fallen gelassen.

2.5 Aussagekraft der Ergebnisse

Bei der Interpretation der Daten der schriftlichen Befragung ist zu berücksichtigen, dass die Stichprobenergebnisse von den realen Werten abweichen können. Der Stichprobenfehler beträgt bei einer Teilgruppe von 100 Befragten maximal $\pm 8.4\%$, bei 400 Befragten max. 4.4% . Beim Vergleich von Teilgruppen empfehlen wir als Faustregel, Abweichun-

¹ Berechnung auf der Basis von $N = 824$ minus $17 = 807$ Zieladressen (Adresse nicht richtig, Doublette, Zielperson im Ruhestand u.a.).

gen von weniger als 10% zwischen zwei Untersuchungsgruppen kein zu grosses Gewicht beizumessen.

Die Ergebnisse der telefonischen Befragung und der offenen Fragen der schriftlichen Befragung haben qualitativen Charakter. Sie sollen die Ergebnisse der quantitativen Befragung illustrieren und insbesondere die Frage der Nutzung und des Nutzens vertiefen. Wo eine Quantifizierung vorgenommen wird, hat dies lediglich den Zweck, die Grösse unterschiedlicher Teilgruppen annäherungsweise zu beschreiben.

Aufgrund der guten Teilnahmequote und der hohen Übereinstimmung und Plausibilität der Aussagen aus schriftlicher und telefonischer Befragung kann davon ausgegangen werden, dass die hier dargestellten Daten eine hohe Aussagekraft haben und für die Weiterentwicklung des Klassencockpit genutzt werden können.

3. Ergebnisse

3.1 Bekanntheit und Verbreitung des Klassencockpit (Repräsentativbefragung)

Klassencockpit verfügt über eine hohe Bekanntheit und hat in der Lehrerschaft des Kantons St. Gallen eine grosse Verbreitung. 85% der Lehrerinnen und Lehrer nutzen Klassencockpit aktuell, 6% kennen das Produkt, haben es aber noch nicht genutzt. Weiteren 6% ist es unbekannt und 3% nutzen es aktuell nicht mehr.

Tabelle 1: Bekanntheit und Nutzung von Klassencockpit

	Dienstalter (auf aktueller Stufe)				Stufe			Total
	bis 4	5-10	11-24	25+	MS	Real	Sek	
bekannt, aber noch nicht genutzt	5.1	5.9	4.9	5.6	1.5	5.7	11.2	5.5
schon selber genutzt, heute nicht mehr	2.5	4.4	2.8	2.8	1.0	3.4	5.6	3.0
aktuelle NutzerIn	75.9	85.3	88.0	90.2	90.5	86.2	79.7	85.5
unbekannt	16.5	4.4	4.2	1.4	7.0	4.6	3.5	6.0
Anzahl Befragte	79	68	142	143	201	87	143	435

Klassencockpit ist auf der Mittelstufe etwas mehr verbreitet als in der Real- beziehungsweise Sekundarschule. Je höher das Dienstalter, desto bekannter ist Klassencockpit und desto mehr wird es genutzt.

58% derjenigen, denen Klassencockpit zwar bekannt ist, die es aber bislang nicht genutzt haben (5.6% der Stichprobe), haben schon erwogen, es einzusetzen. Der Hauptgrund, weshalb es bislang noch nicht dazu kam, waren die mangelnden Einsatzmöglichkeiten. Unter den ehemaligen Nutzerinnen und Nutzern haben 59% einen erneuten Einsatz erwogen.

Reichweite des Klassencockpit

Befragungsergebnisse können nicht *tel quel* auf die Population – alle Lehrpersonen der interessierenden Stufen – hochgerechnet werden, vor allem nicht, wenn die Verbreitung eines Produktes eingeschätzt werden will. Wenn die Teilnahmequote nicht (annähernd) 100% erreicht, müssen Überlegungen darüber angestellt werden, wie sehr sich Stichprobe und Population in Bezug auf interessierende Kriterien unterscheiden. Erfahrungsgemäss sind Alter, Geschlecht, Dienstalter, geografische Verteilung u.a. der Lehrpersonen in Befragungen in praktisch gleichen Proportionen wie in der Grundgesamtheit verteilt. Etwas schwieriger ist es, Aussagen über inhaltliche Kriterien zu machen (hier: Bekanntheit und Nutzung des Klassencockpit).

Wollen also Aussagen über *alle Lehrpersonen* auf den interessierenden Schulstufen gemacht werden, müssen Stichprobenverzerrungen eingeschätzt werden. Aufgrund der Befragungsdaten können dann minimale und maximale Reichweite des Klassencockpit

berechnet beziehungsweise reale Werte geschätzt werden. Hinter entsprechenden Einschätzungen stehen konservative oder optimistische Annahmen. Mit dem Ziel der besseren Lesbarkeit der Ausführungen beschränken wir uns auf Berechnungen für die ganze Stichprobe. In *Tabelle 2* werden Rechenbeispiele dargestellt, die auf *drei unterschiedlichen Annahmen* basieren:

- a) *Repräsentative Stichprobe (keine Verzerrungen)*: Die realisierte Stichprobe bildet die Population der Volksschullehrerinnen und Volksschullehrer der 4. Primarklasse bis zur 3. Oberstufenklasse *verzerrungsfrei* ab. Es wird also angenommen, dass Bekanntheit und Nutzung bei den BefragungsteilnehmerInnen gleich verteilt sind wie bei den eine Teilnahme Verweigernden.
- b) *Mittlere Verzerrung der Stichprobe*: Ein Drittel 33% (Annahme) der an der Befragung nicht teilnehmenden Lehrpersonen kennen und nutzen Klassencockpit nicht und haben aus diesem Grund den Fragebogen nicht zurückgesandt. Die übrigen nicht Teilnehmenden verteilen sich bezüglich Kennen und Nutzen des Klassencockpit gleich wie in der Stichprobe.
- c) *Grosse Verzerrung der Stichprobe*: Die Lehrpersonen, die den Fragebogen nicht ausgefüllt haben, nutzen Klassencockpit nicht.

Aufgrund dieser Annahmen ergeben sich Nutzungsquoten von 47-86% (*s. Tabelle 2*).

Tabelle 2: Schätzwerte für die Bekanntheit und Nutzung von Klassencockpit

	Annahmen zur Repräsentativität der Stichprobe		
	a) repräsentativ	b) mittlere Verzerrung	c) grosse Verzerrung
NutzerInnen	85.5%	77.0%	46.9%
nicht NutzerInnen	14.5%	23.0%	53.1%

Eigene methodische Studien im Rahmen der für den LCH realisierten *Arbeitszeituntersuchung (1999)* und der *Berufszufriedenheitsstudie (2002)* haben geringe Stichprobenverzerrungen ergeben. So wich bei der Arbeitszeiterhebung die selbst geschätzte Arbeitszeit der Kontrollgruppe kaum von derjenigen der Stichprobe ab. Ebenso bei der Berufszufriedenheitsstudie, wo die Ergebnisse auch bei sehr unterschiedlichen Teilnahmequoten in Stichprobe (26%) und Kontrollgruppe (75%) nur minimal voneinander abwichen.

Aufgrund dieser Überlegungen ist es realistisch, heute von einer Reichweite des Klassencockpit im Bereich von 77% bis maximal 86% auszugehen. Die Zahl 80% dürfte realistisch sein.

Einschätzung der Leistungsfähigkeit

Unter sieben zur Beurteilung vorgegebenen Funktionen des Klassencockpit, werden von den Befragten praktisch alle als erfüllt oder mindestens teilweise erfüllt eingeschätzt. Der Anteil der Lehrpersonen, die eine Leistung als gar nicht erfüllt betrachten, schwankt zwischen 2% und 19%.

Klassencockpit leistet gemäss Lehrpersonen vor allem die Funktion, *mit anderen Klassen zu vergleichen* (74% betrachten diese Funktion als vollständig erfüllt). Je 39% sehen die Funktion der *Objektivierung der Beurteilungspraxis* erfüllt bzw. erkennen Klassencockpit als *Selbstevaluationsinstrument für den eigenen Unterricht*, 31% sehen es als *Grundlage für Elterngespräche*. Zwischen 19% und 25% betrachten die anderen Leistungen als erfüllt.

Mit Ausnahme der Leistung *Vergleich mit anderen Klassen* ist die Gruppe der Lehrpersonen, die eine Funktion als teilweise erfüllt betrachten, grösser als die beiden anderen Teilgruppen (erfüllt; nicht erfüllt). Die Bewertung fällt in der Regel um so günstiger aus, je länger eine Lehrperson auf der jetzigen Stufe unterrichtet.

Häufigkeit der Nutzung

Die grosse Mehrheit (83%) nutzt Klassencockpit an *allen* Terminen. 13% entscheiden von Fall zu Fall und 3% nutzen nur einen oder zwei Termine. Zu den besonders intensiven Nutzerinnen und Nutzern gehören die Mittelstufenlehrpersonen (89% nutzen alle Termine; 73% der Sekundarlehrpersonen) und die jüngeren Lehrkräfte (90% der Lehrpersonen mit vier und weniger Praxisjahren). Je länger jemand auf der jetzigen Stufe unterrichtet, desto häufiger entscheidet er/sie sich *fallweise* für eine Teilnahme. Gleiches gilt für Real- bzw. Sekundarlehrpersonen im Vergleich zu den Primarlehrpersonen.

Tabelle 3: Einsatz des Klassencockpit (Nutzungshäufigkeit)

	Dienstalter (auf aktueller Stufe)				Stufe			Total
	bis 4	5-10	11-24	25+	MS	Real	Sek	
an allen Terminen	90.0	81.0	85.6	77.5	88.5	84.0	72.8	83.7
entscheide von Fall zu Fall	8.3	12.1	12.0	16.3	6.0	13.3	23.7	12.9
nutze nur einen bzw. zwei Termine	1.7	1.7	1.6	6.2	3.8	1.3	3.5	3.3
keine Angabe		5.2	0.8		1.6	1.3		1.1
Anzahl	60	58	125	129	182	75	104	372

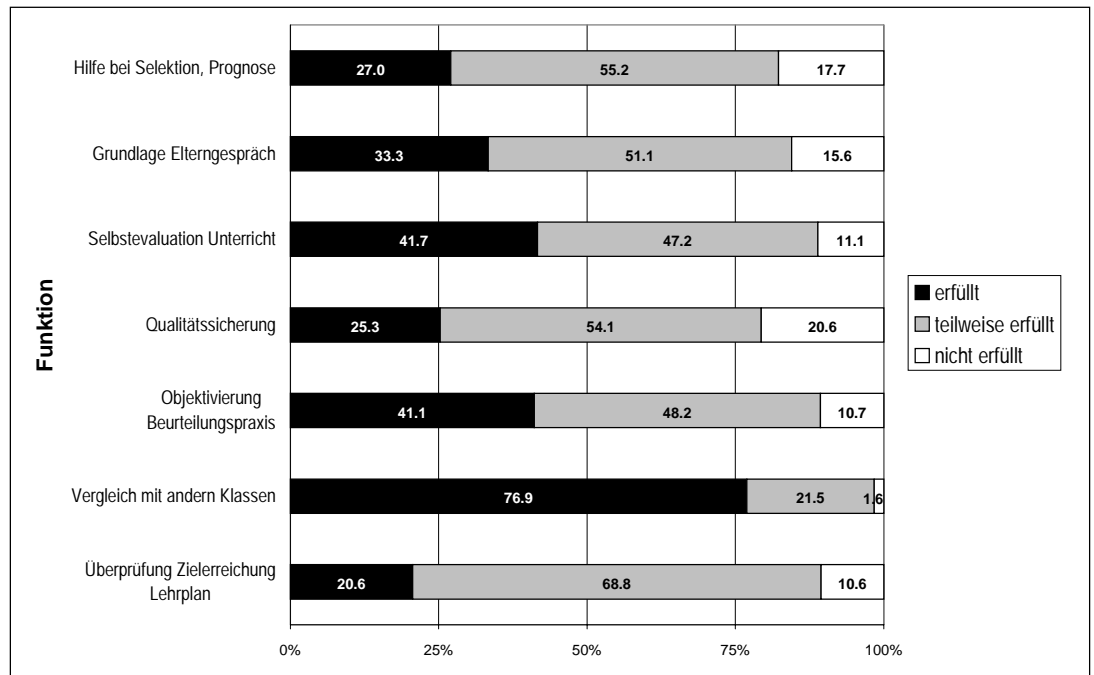


Abbildung 1: Einschätzung der Leistungsfähigkeit von Klassencockpit

Wie aus *Abbildung 2* ersichtlich, unterscheiden sich die Beurteilungen der einzelnen Schulstufen nur in Nuancen. Der deutlichste Unterschied ist feststellbar zwischen den Sekundarlehrpersonen und den übrigen Stufen in Bezug auf Klassencockpit als *Hilfsmittel bei Selektionsentscheiden und als Prognoseinstrument* sowie als *Grundlage für Elterngespräche*.

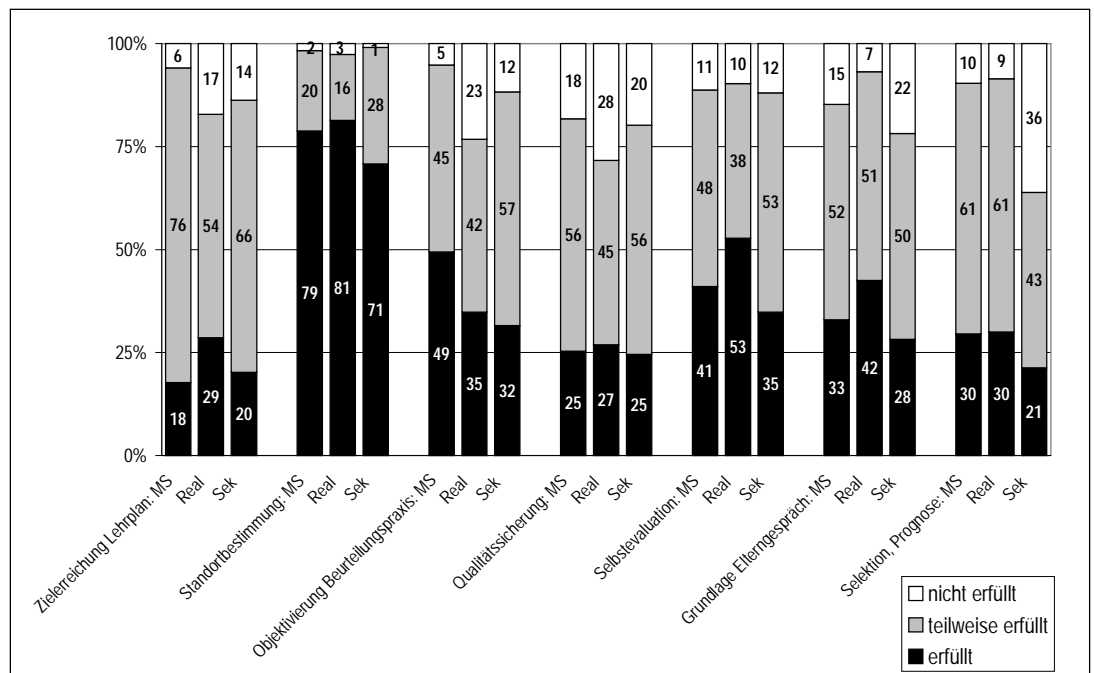


Abbildung 2: Bedeutung von Klassencockpit nach Schulstufe

3.2 Anwendung des Klassencockpit

Die aktuelle Klassencockpit-Nutzerschaft ist zu einem Fünftel seit der Pilotphase dabei. Knapp 40% der Lehrpersonen nutzen Klassencockpit seit 3 Jahren oder länger.

Tabelle 4: Einstieg in die Nutzung von Klassencockpit

Jahr	Anzahl	in %
1995-97	67	18.0
1998	43	11.6
1999	63	16.9
2000	79	21.2
2001	63	16.9
2002	47	12.6
Keine Angabe	10	2.7
Total Befragte	372	100.0

Bedeutung von Klassencockpit im Unterricht

Um den Einsatz des Klassencockpit besser verstehen zu können, wurde nicht nur die Leistungsfähigkeit abgefragt, sondern auch die Bedeutung nach Verwendungszweck. Mit den entsprechenden Angaben können die zur Leistungsfähigkeit gemachten Angaben relativiert beziehungsweise geeicht werden.

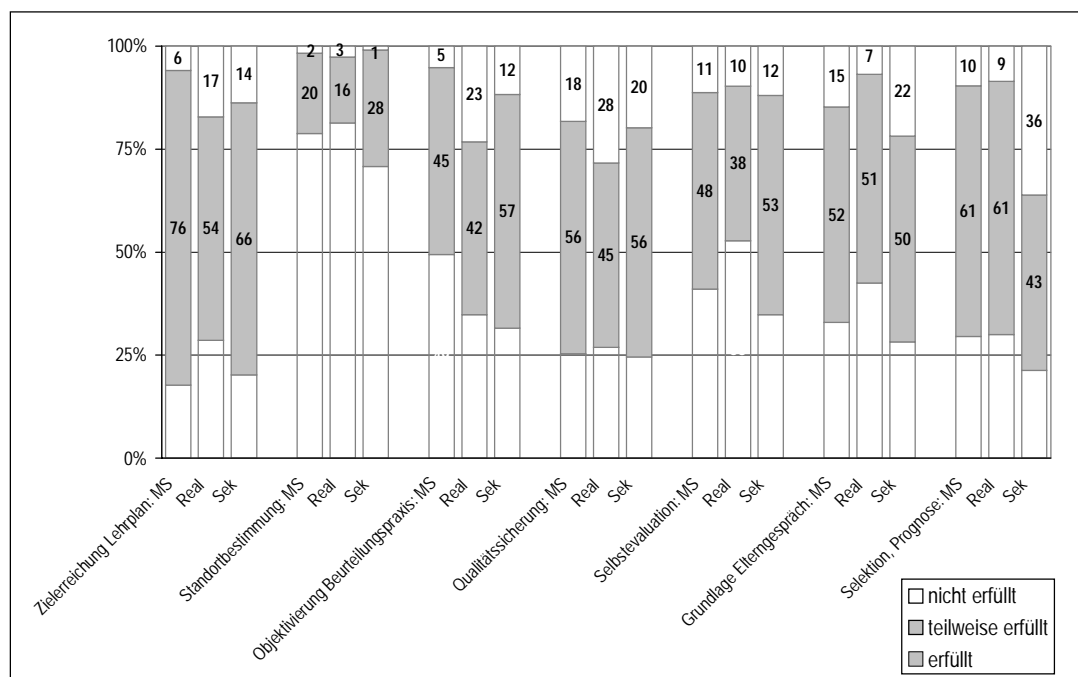


Abbildung 3: Bedeutung von Klassencockpit im Unterricht

Allgemein ist die Bedeutung des Klassencockpit bei der Planung und Führung des Unterrichts etwas geringer, als dies aufgrund der Beurteilung der Leistungsfähigkeit hätte erwartet werden können. Gemessen an der Häufigkeit des Höchstwertes (grosse Bedeutung) hat Klassencockpit klar die grösste Bedeutung (bei 53% der Lehrpersonen) als Instrument für den Vergleich der eigenen Klasse mit anderen. Relativ grosse Bedeutung hat er für 30% zur Objektivierung der Beurteilungspraxis und für 28% die Selbstevaluation des Unterrichts.

Vergleicht man die Stufen, fallen die Sekundarlehrpersonen als tendenziell kritischer auf als die übrigen, die Reallehrpersonen als tendenziell positivste. Dieser Befund muss wohl auch mit Blick auf die unterschiedlich lange Erfahrung mit Klassencockpit etwas relativiert werden.

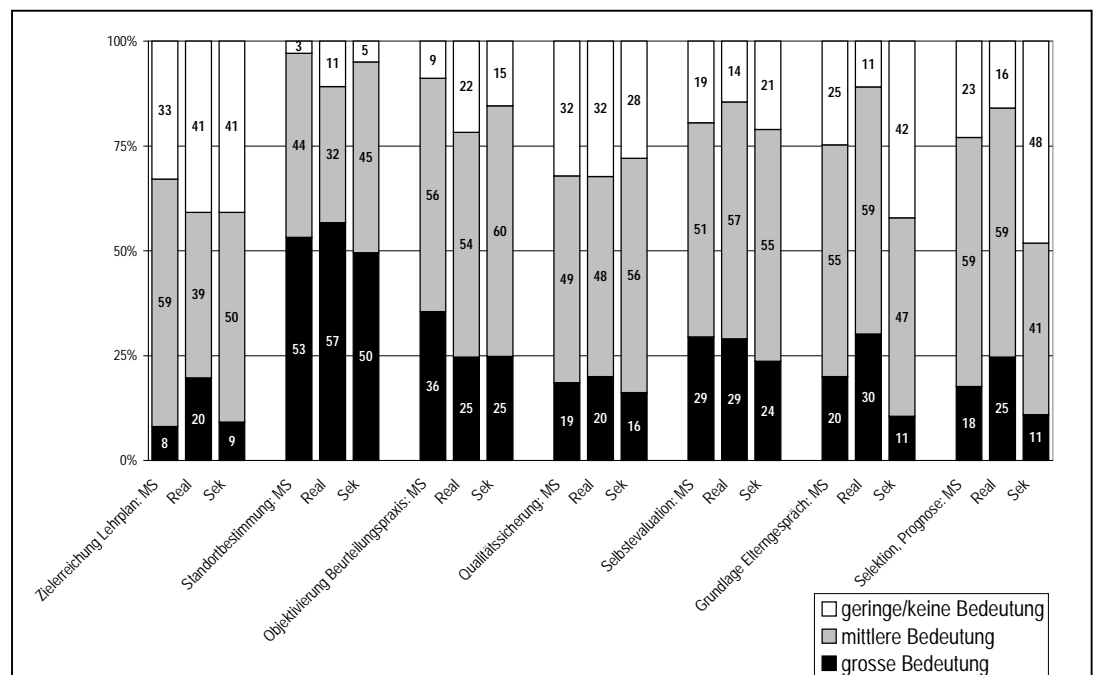
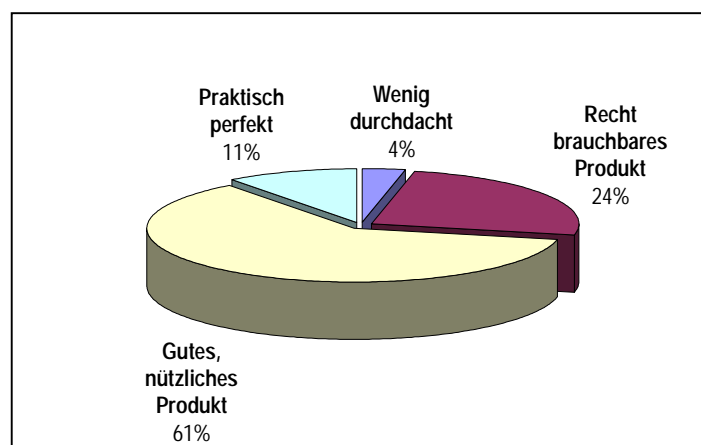


Abbildung 4: Einschätzung der Leistungsfähigkeit nach Schulstufe

Abbildung 5: Beurteilung des Klassencockpit



Die Beurteilung von Klassencockpit fällt eindeutig aus. Auf einer 5-stufigen kategorialen Skala wählen 61% den zweithöchsten (gutes, nützliches Produkt mit einigem Verbesserungspotenzial) und 10% den höchsten Wert (sehr gutes, nützliches Produkt, praktisch perfekt). Der tiefste

Wert (*unbrauchbares Produkt*) wurde von niemandem gewählt, der zweittiefste (*wenig durchdacht, müsste sehr stark verbessert/überarbeitet werden*) von lediglich 4% aller Befragten.

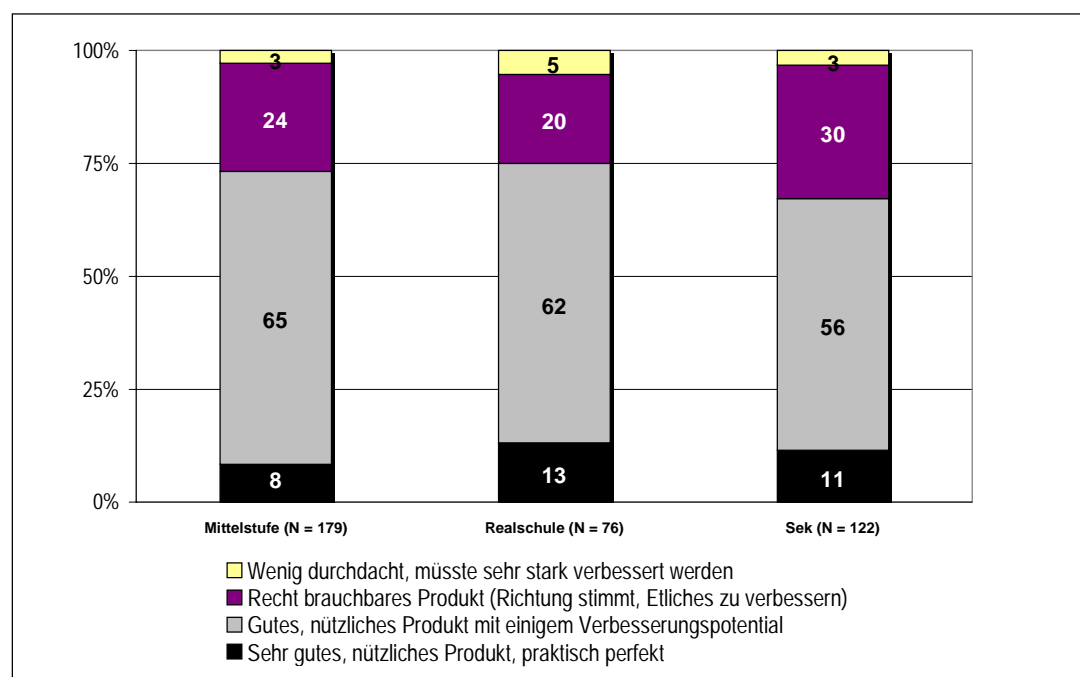


Abbildung 6: Beurteilung des Klassencockpit nach Schulstufe der Lehrpersonen

Die Einschätzung des persönlichen Nutzens durch die diversen Stufen ist ziemlich ausgeglichen. Sekundarlehrpersonen sind tendenziell etwas skeptischer im Urteil als ihre Kolleginnen und Kollegen der beiden anderen Stufen.

Um die Beurteilung des Klassencockpit noch etwas genauer anzuschauen, haben wir die Stichprobe in drei Kontrastgruppen gemäss folgender Matrix unterteilt:

Tabelle 5: Matrix zum Nutzen des Klassencockpit

		Bedeutung der Funktion für Unterrichtsplanung/-führung		
		gering	mittel	gross
Funktion erfüllt	nein			
	teilweise			
	ja			

geringer Nutzen
 mittlerer Nutzen
 grosser Nutzen

Die kombinierte Auszählung der beiden Fragen Auszählung der kombinierten Antwortmöglichkeiten ergab folgendes Ergebnis:

Tabelle 6: Nutzen nach Teilfunktionen Klassencockpit (Stichprobenanteile in %)

Teilfunktion	Nutzen gemäss Matrix Tabelle 5 / Stichprobenanteile		
	gross	mittel	gering
Standortbestimmung Klasse	76.0	19.9	4.1
Objektivierung Beurteilungspraxis	44.8	40.7	14.5
Selbstevaluationsinstrument	42.1	40.4	17.5
Grundlage für Elterngespräch	31.7	40.6	27.7
Hilfe für Selektion und Prognose	26.6	43.6	29.8
Qualitätssicherung	26.4	40.8	32.7
Überprüfung Zielerreichung	21.9	42.9	35.2

Gemäss dieser Datenauswertung dürfte für die Lehrpersonen auf der Ebene *Standortbestimmung im Vergleich zu anderen Klassen* der grösste Nutzen anfallen. Auch in Bezug auf *Objektivierung der Beurteilungspraxis* und *Selbstevaluationsinstrument* kann ein angemessener Nutzen festgestellt werden. Zumeist über 40% der Stichprobe gehören zur Gruppe, die den Teilfunktionen des Klassencockpit einen mittleren Nutzen abgewinnen können. Die Gruppe derer, die in Bezug auf eine Funktion überhaupt keinen oder nur geringen Nutzen aufweisen dürfte, erreicht eher tiefe Stichprobenanteile von 4% bis maximal 35%.

Der Vergleich der befragten Nutzerinnen nach dem Kriterium der Nutzungsdauer zeigt, dass für Lehrpersonen die Bedeutung von Klassencockpit mit wachsender Anwendungsdauer (Erfahrung) wächst (*Tabelle 7*).

Tabelle 7: Teilfunktionen mit grosser Bedeutung nach Anwendungsdauer (%-Anteile)

Teilfunktion	Klassencockpit hat grosse Bedeutung/wenden Klassencockpit an seit:		
	1995	1998	2001
Standortbestimmung Klasse	79.3	88.4	66.7
Objektivierung Beurteilungspraxis	39.7	46.5	34.9
Selbstevaluationsinstrument	41.4	41.9	33.3
Grundlage für Elterngespräch	43.1	39.5	25.4
Hilfe für Selektion und Prognose	31.0	30.2	15.9
Qualitätssicherung	34.5	27.9	17.5
Überprüfung Zielerreichung	36.2	16.3	12.7

3.3 Qualitative Ergebnisse

Die Ergebnisse der Auswertung der *qualitativen Gespräche* und der *offenen Fragen* der schriftlichen Befragung decken sich praktisch vollständig. Aus diesem Grund sowie mit dem Ziel der Leserfreundlichkeit werden die beiden Datenquellen zusammen dargestellt und wird nur ausnahmsweise auf die Quelle der Aussage verwiesen.

Übersicht über die häufigsten Inhalte der Antworten auf offene Fragen

In 331 Fragebogen wurden Antworten auf offene Fragen registriert. Diese verteilten sich auf verschiedene Fragen, wurden aber als ein einziger Datensatz behandelt, da sich die vorgenommenen Einträge nicht konsequent auf die jeweilige Frage bezogen. Die Antworten wurden nach einzelnen, kumuliert 473 Bedeutungseinheiten aufgeschlüsselt und anschliessend übergeordneten Inhaltskategorien zugewiesen.

Wir definierten folgende groben Inhaltskategorien:

Tabelle 8: In offenen Fragen vorgefundene Inhaltskategorien

Inhaltskategorie		Anzahl	in %
Zufriedenheit	allgemein, Einzelaspekte	55	11.6
	Auswertung	16	3.4
	Schülerreaktionen	8	1.7
Kritik, Skepsis	Noten, Bewertungsansatz	108	22.8
	Aufgaben	73	15.4
	Lehrplan- und Lehrmittelkompatibilität	62	13.1
	unspezifisch, marginale Einzelaspekte	50	11.4
	Missbrauch des Instruments	19	4.0
	Auswertung	18	3.8
	Stufenmodul	15	3.2
	Schülerreaktionen	13	2.7
Wünsche, Vorschläge		36	7.6
Total Bedeutungseinheiten		473	100.0

Wie in praktisch allen schriftlichen Befragungen, werden bei der Beantwortung der offenen Fragen kritische Standpunkte häufiger notiert als positive Rückmeldungen. In der qualitativen Telefonbefragung waren Komplimente für den Klassencockpit und Kritik in etwa ausgewogen, die Tonalität war aber allgemein beifällig gehalten.

Die Antworthäufigkeiten konzentrieren sich auf drei Kategorien, welche zusammen mehr als 50% aller ermittelten Bedeutungseinheiten einschliessen: *Noten/Bewertungsansatz*, *Aufgabenstellung* und *Lehrplan- bzw. Lehrmittelkompatibilität*.

Einsatz des Klassencockpit im Unterricht

Die Anwendung des Klassencockpit ist in der Regel mehrfacher Art. Im Vordergrund steht die *Standortbestimmung* für die Klasse, d.h. der Vergleich mit anderen Klassen derselben Schule oder – insbesondere auch für Lehrpersonen ohne KollegInnen mit gleicher Klasse – des Kantons. Damit verbunden wird eine Verortung der eigenen (Lehrer)Leistung in einem grösseren Zusammenhang (d.h. gegenüber anderen Lehrpersonen). Der Wunsch, Sicherheit über seine Position durch den Beizug externer, objektiverer Tests zu gewinnen, ist in zahlreichen mündlichen und schriftlichen Äusserungen erkenn-

bar. Mehrmals wird das Bild vom «Einzelkämpfer» gezeichnet, der auf verlässliche Rückmeldungen angewiesen ist.

Wahrnehmung des Klassencockpit und Anwendung

Mehrere Lehrpersonen erwarten vom Klassencockpit offenbar eine Funktion, die gerade nicht geleistet werden kann: gewissermassen eine individuelle Rückmeldung auf den jeweiligen Stand der Klasse – wobei *die Klasse, ihre Lehrperson und der Stand der behandelten Inhalte selber als Referenzrahmen dienen sollen*. Das mehrmals gezeichnete Bild von „zu guten“ oder „zu strengen Benotungen“² interpretieren wir als Beleg für dieses offensichtliche Missverständnis. Es scheint vor allem dann zu aufzukommen, wenn die im Klassencockpit-Test von einer Klasse erbrachten Leistungen nicht den Erwartungen der Lehrperson entsprechen. Die Kernidee von Klassencockpit – einen in der (für den Kanton repräsentativen) Eichstichprobe realisierten Notenschlüssels auf alle teilnehmenden Klassen zu transformieren und dadurch Orientierung zu ermöglichen – wird genau in dem Moment ausgeblendet, wo sie zum Tragen kommt.

Die hier angesprochene Gruppe von Lehrpersonen ist allerdings die deutlich kleinere. Die andere ist sich der Möglichkeit von grossen Abweichungen bewusst und sucht zu ergründen, wie sehr die gestellten Aufgaben (und die Gewöhnung an sie), die Vergleichsstichprobe,³ der Stand der Arbeit bezogen auf Lehrplan oder Lehrmittel, die eigene Schülerpopulation oder die momentanen Schülerleistungen Grund für die Ergebnisse in einem Modul waren. Wir hatten in den Befragungen den Eindruck, dass Lehrpersonen, die sich sorgfältig diese Überlegungen machen, nicht nur ein realistisches (richtiges) Bild von Klassencockpit hatten, sondern die Teilnahme auch als fachlich spannende Arbeit empfanden. So bezeichnete eine Lehrperson Klassencockpit lobend als *„echten wissenschaftlichen Vergleich ohne Wenn und Aber.“*

Aufgrund dieses Zusammenhangs ist auch einsichtig, dass die getesteten Inhalte der Aufgaben unterschiedlich charakterisiert wurden. Lobten die einen etwa die *„weiten und aussagekräftigen Problemstellungen“*, die *„andere Form von Prüfungsaufgaben“* oder fanden Klassencockpit *„modern, fantasievoll, nicht lernbar, objektiv und zukunftsorientiert“*, so kritisieren andere das Fehlen von Lerninhalten (so etwa *„knallharte Grammatik statt Bildverständnis“*, *„Arbeiten zum Thema Texte schreiben“*, *„Geometrieinhalte“* usw.) oder die *„Einseitigkeit“* einzelner Module, d.h. dass *„nur Teilaspekte, z.B. Hörverständnis“* geprüft werde. *„Aspekte wie Kreativität, Vorstellungsreichtum, Sensibilität, Beharrlichkeit [...] kommen in allen mir bekannten Modulen eindeutig zu kurz.“* Besonders deutlich wird der Einwand in Bezug auf inhaltliche Lücken oder Einseitigkeit von Lehrpersonen geltend gemacht, die Klassencockpit für die Selektion einsetzen möchten. Ein Befragter äussert sich sehr direkt wie folgt: *„Da meine SchülerInnen gewisse Aufgaben noch nicht lösen können, kann ich die Note nicht im Zeugnis verwenden.“* Aufgrund von

² In Anführungszeichen und kursiv abgesetzte Textteile beinhalten Zitate aus den qualitativen Gesprächen oder den Fragebogen.

³ Offenbar misstrauen einige Lehrpersonen der Eichstichprobe. So wird gewünscht, aktuelle Resultate laufend (*„nach Klasse und Region“*) in die Mittelwertsberechnungen aufzunehmen, *„so dass Anpassungen (im Schnitt) vorgenommen werden können.“*

Aussagen von Befragten ist davon auszugehen, dass einzelne Lehrpersonen die Module abgleichen mit ihren Lehrinhalten und nicht teilnehmen, wenn eine Diskrepanz festgestellt wird: „*Letztes Jahr machte ich bei einem Modul mit, dieses Jahr war die Stoffauswahl einseitig.*“ Ein Befragter lädt die Projektleitung ein, angesichts der inhaltlich begrenzten Testmöglichkeiten bescheiden zu bleiben.

Mehrere Male kommt von Befragten der (zumeist auf das „*Flughafen-Modul*“ bezogene) Hinweis, die Aufgaben stammten nicht aus der Lebenswelt ihrer Schülerinnen und Schüler. Kritisiert wurde auch die Verknüpfung von Deutsch (oder Mathematik) mit Testinhalten, die „*nichts mit diesen Fachbereichen*“ zu tun hätten: Fahrplan lesen/Deutsch, Mathematik/Sprachverständnis u.a.

Nicht von allen verstanden wird die Logik der Versuchsgruppe, die zur Test-Eichung benutzt wird. So wird bisweilen davon ausgegangen, ein Jahrgangsvergleich sei erst möglich, wenn alle SchülerInnen im Kanton erfasst werden; die Bildung von Vergleichswerten mittels Stichproben ist nicht allen geläufig. Allerdings wird auch von Lehrpersonen, die hier keine Verständnisprobleme haben, angemerkt, die Eichstichprobe sei mit 25 Klassen etwas klein.

Objektivierte Schülerbeurteilung

Ähnlich dem Wunsch nach einer Einordnung der Klasse in einer grösseren Schülerpopulation, will auch die vorgenommene Schülerbeurteilung überprüft werden. Die in den qualitativen Gesprächen gemachten Aussagen zu diesem Punkt zeigen, dass sich diejenigen Lehrpersonen schwer tun, die noch über wenige Erfahrung mit Klassencockpit verfügen. Die Anwendung mehrerer Module bewirkt ein besseres Verständnis des Instruments, können doch die generellen Schwankungen oder einzelnen Ausreisser von SchülerInnen besser interpretiert werden.

Allgemein wird die Idee, die eigene Beurteilungspraxis zu hinterfragen, begrüsst. In der Praxis klaffen Werte des Klassencockpit und eigene Erfahrung oft so sehr auseinander, dass die Qualität von Klassencockpit hinterfragt wird – immer ausgehend von der Vorstellung, eigene und mit Klassencockpit gemachte Erfahrungen müssten sich einiger massen decken. Ein mehrfach gemachter Hinweis bezieht sich auf das Problem, dass der Lehrplan nicht chronologisch gegliedert ist und deshalb je Klasse zum Zeitpunkt der Klassencockpit-Erhebung unterschiedliche Inhalte behandelt worden sind (*vgl. dazu auch Zielerreichung Lehrplan weiter unten*).

Gestaltung/Aufmachung/Ausfertigung des Klassencockpit

Die äussere Aufmachung und Gestaltung von Klassencockpit findet mehrheitlich grossen Gefallen: „*Attraktive Gestaltung*“ ist denn auch ein häufiges Urteil in den telefonischen Gesprächen. Einige wünschen sich eine etwas schlankere Ausführung. Jeweils von einzelnen Lehrerinnen oder Lehrern wird „*die Papierflut*“ bemängelt. Im grünen Beiblatt werden wieder Auswertungs- und Notentabellen gewünscht.

Der Einsatz von Internet wird allgemein gelobt, wenngleich bisweilen auch auf Lehrpersonen aufmerksam gemacht wird, die „*noch nicht ans Internet*“ gewohnt seien. Der In-

ternet-Auftritt selber findet hie und da negative Bemerkungen wegen gestalterischen und Benutzeraspekten, so etwa wegen zu viel Animation, zu kleiner Schrift, Eingabemodus (gewünscht wird Tastenbefehl statt Mausklick bei Multiple Choice-Aufgaben oder bei der Eingabe der Punktwerte) und nicht ausdrückbarer Notenschlüssel.

Bedienung/Auswertung

Die Mehrzahl der Befragten dürfte mit der Bedienung von Klassencockpit keine Probleme haben. Geschätzt werden vor allem die einfache Korrekturarbeit und die Internetauswertung inkl. verwendete statistische Parameter. Einzelne AnwenderInnen nehmen jedoch kritische Anmerkungen vor, wo die meisten Antwortenden keine Probleme haben, sowohl allgemein („*die Bedienung könnte freundlicher sein*“) wie auch spezifisch (Ausdruck, erzwungene Internetnutzung, unbefriedigende Korrekturhilfsmittel, Materialaufwand).

Selektion, Prognose

Klassencockpit wird als zusätzliche Information beigezogen. Die Verwendung von individuellen oder Klassenergebnissen gegen aussen erfolgt relativ diskret. Für die Unterstützung von Promotionen wird Klassencockpit praktisch nicht verwendet, wohl aber als zusätzlicher (negativer) Beleg um Eltern zu überzeugen, dass ein Kind oder ein Jugendlicher nicht die vermeintliche Leistung erbringt.

Sowohl in den Fragebogen als auch in einzelnen Gesprächen wird aber ersichtlich, dass die Lehrpersonen gerne mehr hätten. Das Interesse, über ein unbestechliches (und entlastendes) Instrument in Selektionsprozessen, über «harte Fakten» zu den Schülerleistungen zu verfügen, ist unübersehbar. Allerdings scheint sich bei den meisten die Einsicht durchgesetzt zu haben, dass Klassencockpit diese hohe Anforderung nicht erfüllen kann. Aus diesem Grund wird denn auch gewünscht, dass im Leistungsbeschrieb entsprechend sorgfältige Formulierungen gewählt werden.

Elterngespräche

Für *Elterngespräche* wird Klassencockpit in mittlerer Häufigkeit und oft mit dem gleichen Ziel eingesetzt: Ein (eher) schlechtes Bild von den Leistungen eines Schülers oder einer Schülerin wird durch die Ergebnisse im Klassencockpit ergänzt, was gemäss Aussagen mehrerer GesprächspartnerInnen „überzeugend“ wirkt. Die Lehrerinnen und Lehrer sehen sich „*aus der Schusslinie*“ genommen, wenn sie die Testresultate vorlegen können. Im Gespräch können *die Leistungen* der SchülerInnen fokussiert werden, die explizit oder implizit oft im Zentrum der Kritik stehende Lehrperson ist nurmehr mittelbar involviert.

Es gibt nicht wenige Lehrpersonen, die Klassencockpit entweder nicht oder sehr gezielt einsetzen. So werden einem Teil vor allem der fremdsprachigen Eltern *nur* die Noten gezeigt, weil diese sich für die interessanten Feinheiten dahinter (wo muss Kind lernen, wo sind keine Probleme) kaum interessieren. Andere Lehrpersonen behalten gerade aus diesem Grund die Noten von Klassencockpit für sich.

Die qualitativen Gespräche führen uns zur Hypothese, dass der Einsatz von Klassencockpit im Elterngespräch nicht ganz einfach ist. Zum einen ist der Kontakt mit Eltern für Lehrpersonen allgemein etwas vom Schwierigeren des Berufsalltags. Entsprechende Gespräche zudem mit einem Instrument zu verbinden, das noch nicht ganz beherrscht und/oder durchschaut wird, wird dann auch als besonders problematisch empfunden.

Wir verweisen aber auch auf die *Tabelle 7* und erinnern daran, dass Klassencockpit gerade im Bereich Elterngespräch von einem ansehnlichen Anteil derjenigen Lehrpersonen grosse Bedeutung beigemessen wird, die das Instrument relativ lang einsetzen.

Qualitätssicherung/Selbstevaluation

Vor allem in den telefonischen Gesprächen haben wir bei etlichen Befragten ein hohes Engagement herausgespürt im Bemühen um eine gute Qualität des Unterrichts. Für diese Lehrpersonen kommt Klassencockpit als Zusatzinstrument gelegen, suchen sie doch in der Planung und Auswertung des Unterrichts verschiedene Instrumente einzusetzen. Aus einigen Aussagen spricht eine gewisse Enttäuschung über Lehrmittel allgemein, die nicht in der Lage seien, *verschiedene Fähigkeiten* in Kombination miteinander zu testen. Wir vermuten, dass viele Lehrpersonen, die sich im schriftlichen Fragebogen begeistert über die Testaufgaben ausgesprochen haben, gerade diese Besonderheit des Klassencockpit schätzen.

Andererseits ist auch eine gewisse Kluft festzustellen zwischen den Vorstellungen dieser Lehrpersonen und den Testmodulen. Viele haben sich inzwischen angewöhnt, Schülerinnen und Schüler auch für falsche Ergebnisse, aber richtige Lösungswege zu belohnen. Die Module wirken hier in den Augen der Befragten etwas starr; die Bemühungen, die Sache mit halben oder Viertelpunkten wieder ins Lot zu bringen, sind recht verbreitet. Mit der Art der Aufgabenstellung und der Tatsache, dass einiges, was getestet wird, möglicherweise im Unterricht noch gar nicht behandelt wurde, können nicht alle Lehrpersonen gleich gut umgehen. Einige lassen sich bisweilen schnell dazu verführen, das Ergebnis zu relativieren: „*Da schlage ich jeweils nicht gleich Alarm.*“ In der Praxis scheiden sich denn auch die Geister, wenn es bei schlechter Klassenleistung darum geht, die Ursachen nicht nur bei den Aufgaben, sondern auch bei sich selber zu suchen.

Ergebnisse von Tests werden in einigen Schulen über Jahre verglichen. Dies wird insofern positiv gewertet, als damit sich gewissermassen ein «Schulgedächtnis» bilden kann, das die Ergebnisse früherer Jahrgänge systematisch speichert.

Zielerreichung Lehrplan

Lehrplan/Lehrmittel sind die Stichwörter, die den Befragten auf der Zunge brennen. Der Mehrzahl scheint bekannt zu sein, dass die Schwierigkeiten breit angelegter Vergleichstests in den Lehrplänen angelegt sind. Da Stufenlernziele Jahrgangsziele abgelöst haben bewegen sich die Lehrpersonen in den meisten Fächern relativ frei im Rahmen der vorgegebenen Zielsetzungen. Anders als in alten Lehrplankonzepten variiert dadurch natürlich die chronologische Abfolge der gelehrten Inhalte. Bezeichnend für den verbreiteten Wunsch von Lehrpersonen, verlässlichere Referenzpunkte zu haben, ist die geringe Akzeptanz der Stufenmodule, die kaum Neugier wecken, wieviel schon gelernt wurde, son-

dern die Lehrpersonen vor allem auf die Frustration jüngerer SchülerInnen hinweisen lassen, was letztlich allerdings „kontraproduktiv“ sei.

Allgemein führt die Art des Lehrplans also dazu, dass zwischen realem Unterricht und Cockpit-Tests nur eingeschränkte Kongruenz zu erwarten ist. Den Lehrpersonen ist nicht immer, aber zumeist klar, dass die Tests auf Lehrplanzielen aufbauen und auf welchen. Dennoch: Ihr Orientierungsinstrument ist vielfach das Lehrmittel. So kann es schon einmal vorkommen, dass sie inhaltlich sich an anderem Ort bewegen als das Testmodul. Wer die Hintergründe dieser Situation erkennt, hat kein Problem und kann die Ergebnisse korrekt verorten. Wer glaubt, es müsse eine Übereinstimmung geben, wird Probleme haben. Wir haben in den schriftlichen Fragebogen genügend Belege für die Existenz beider Gruppen erhalten. Implizit ist der Wunsch da, über mehr Module zu verfügen, wächst doch damit die Wahrscheinlichkeit, Tests über Lehrinhalte zu erhalten, die die bereits anvisierten Lernziele thematisieren.

Dysfunktionale Anwendungen und Angst vor Missbrauch

Verschiedene Befragte weisen auf Unarten der Anwendung hin:

- Einige Kollegen *üben* für Klassencockpit
- Klassencockpit wird als Aufgabe nach Hause abgegeben
- Klassencockpit wird als Prüfung eingesetzt, die Noten im Zeugnis berücksichtigt, zum Teil lediglich „zum Runden“
- SchülerInnen kommen (zum Teil via LehrerIn) an die Testresultate heran und geben sie weiter
- SchülerInnen *probieren* bei Multiple Choice-Methoden und kommen unverdient zu besseren Ergebnissen.

Diese Punkte rufen Skepsis betreffend die Verlässlichkeit der Resultate hervor und lassen Befragte die Forderung nach unabhängigen Dritten als Tester aufwerfen.

Es gibt aber auch Befürchtungen vor einem Missbrauch des Instruments gegenüber den Lehrpersonen. Ein am Telefon befragter Sekundarlehrer fand Klassencockpit in seinem Fach mit der Bemerkung des Schulleiters, „*er müsse den Test im Mai machen.*“ Anderswo werden ganze Schulen von den Behörden verpflichtet, Klassencockpit einzusetzen. In den Fällen, wo zwischen dem Schulrat und den Lehrpersonen Spannungen bestehen, werden solche Anwendungen natürlich argwöhnisch betrachtet. So wird befürchtet, zunehmend würden Lehrpersonen mittels Klassencockpit gewissermassen mitgetestet.

4. Zusammenfassung, Diskussion und Folgerungen

Um die Bekanntheit, Verbreitung und Nutzung von Klassencockpit zu evaluieren, wurden im Mai 2002 zwei Erhebungen bei der St. Galler Lehrerschaft durchgeführt. Befragt wurden 435 Lehrpersonen mit einem schriftlichen Fragebogen und 38 Nutzerinnen und Nutzer des Klassencockpit mittels eines leitfadengestützten telefonischen Gesprächs. Weiter wurde auszugsweise der E-Mailverkehr zwischen der Projektleitung und anfragenden Lehrpersonen in der Auswertung beigezogen.

Klassencockpit erweist sich als ein von den Lehrpersonen gut akzeptiertes Test-Instrument zur Orientierung über die Schülerleistungen und Selbstevaluation des Deutsch- und Mathematikunterrichts in. Die Beurteilung von Klassencockpit fällt eindeutig positiv aus. Gleichzeitig zeigen sich aber auch einige Problembereiche.

Auf einer 5-stufigen kategorialen Skala wählten 61% den zweithöchsten (*gutes, nützliches Produkt mit einigem Verbesserungspotenzial*) und 10% den höchsten Wert (*sehr gutes, nützliches Produkt, praktisch perfekt*). Der tiefste Wert (*unbrauchbares Produkt*) wurde von niemandem, der zweittiefste (*wenig durchdacht, müsste sehr stark verbessert/überarbeitet werden*) von lediglich 4% aller Befragten gewählt.

Das Instrument verfügt bereits wenige Jahre nach seiner Einführung über eine sehr grosse Reichweite. Fast 95% der *Befragten* kennen Klassencockpit, gegen 86% der Befragten nutzen es bzw. – sehr selten – haben es schon genutzt. Aufgrund von Berechnungen und Schätzungen kann davon ausgegangen werden, dass im Kanton St. Gallen 80% der *Zielgruppe* Klassencockpit nutzen - in der Regel alle verfügbaren Module der entsprechenden Stufe. Hohe Akzeptanz und Nutzungsintensität haben ihre Gründe:

Bedürfnis und Bedarf ausgewiesen

Aus den benutzten Datenquellen, hier insbesondere den telefonischen Gesprächen, wird ersichtlich, dass Lehrpersonen für sich selber – etwa zur Motivierung oder zur Verortung der eigenen Arbeit – auf Informationen von aussen angewiesen sind. Trotz innerer Schulreform (vgl. Teamarbeit, kollegiales Feedback u.a.) immer noch in hohem Masse auf die Selbsteinschätzung angewiesen, suchen sie nach Möglichkeiten und praktischen Instrumenten, um ihren Erfolg und damit auch den Stand der SchülerInnen (oder umgekehrt) zu testen. Die Stichworte dazu: Unterrichtsführung, Unterrichtserfolg, Leistungsstand einzelner SchülerInnen oder der Klasse, Realisierung der Ziele gemäss verwendeten Lehrmitteln bzw. im Lehrplan ausgewiesenen Lernzielen sowie Belegmaterial zur Verwendung gegen aussen. Klassencockpit kann hier sehr viel leisten. Und auf die entsprechenden Angebote sind die Lehrpersonen auch eingestiegen. Je länger und damit besser sie Klassencockpit kennen, desto positiver sind die Rückmeldungen. Dass diese Aussage eine Folge von durch „AussteigerInnen“ bedingten Stichprobenverzerrungen sein könnte, kann mit Blick auf die hohe Reichweite von Klassencockpit nicht bestätigt werden.

Wunsch und Realität sind nicht immer deckungsgleich

Klassencockpit kann aber nicht alles leisten. Einer adäquaten, korrekten Nutzung steht bisweilen der Wunsch der Lehrpersonen entgegen, über ein zuverlässiges Instrument zu verfügen, das den selber getroffenen Selektionsentscheid abstützt oder – noch besser – unanfechtbar macht. Sie suchen nach dem idealen Instrument, das *ihr* Unterrichtsprogramm abfragt und das auch die in der eigenen Klasse, im Schulhaus abgesprochene oder in einer Schulgemeinde etablierte Notengebungskultur berücksichtigt. Schliesslich soll das Messinstrument gemäss diesen Erwartungen auch kontextbezogen (Region, sozio demografische Struktur der Klasse) geeicht sein.

Solche Maximalwünsche erfüllt Klassencockpit nicht, kann Klassencockpit nicht erfüllen. Wenngleich die Nutzerinnen und Nutzer die Teilfunktionen nur zu einem beschränkten Teil *vollständig erfüllt* sehen (Ausnahme: Vergleich mit anderen Klassen), wird die *Nichterfüllung* der Teilfunktionen nur von einer jeweils kleinen Gruppe beklagt. Somit anerkennt der grösste Teil der Lehrpersonen zumindest die teilweise Erfüllung der Funktionen durch Klassencockpit. Diese Einschätzung beziehungsweise Beurteilung steht im Zusammenhang mit der Tatsache, dass der Lehrplan eine punkt- und inhaltgenaue Abfrage von Wissen und Fertigkeiten nicht zulässt, sie hat also mit Klassencockpit zunächst nur mittelbar etwas zu tun. Vor allem auf diesem Hintergrund eröffnet sich indessen ein Feld für Kritik und zum Teil auch Skepsis gegenüber Klassencockpit. Ein häufiger Einwand ist jener, wonach Klassencockpit nur bestimmte Bereiche des Leistungsspektrums erfassen könne (was im Beschrieb allerdings auch so dargestellt ist).

Information und Kommunikation aufrechterhalten

Dass die Liste der Kritik und Wünsche lang ist, kann – mit Blick auf die allgemein gute Beurteilung von Klassencockpit – als Zeichen für das grosse Interesse der Lehrpersonen an der Weiterentwicklung von Klassencockpit verstanden werden. Dieses Interesse manifestiert sich auch im E-Mail-Kontakt von NutzerInnen mit der Projektleitung und war in den Gesprächen mit Lehrpersonen unüberhörbar. Die Präsenz und die am Produktnutzen orientierten Handeln des Projektteams und weiterer Beteiligter werden erkannt und gewürdigt, ebenso wie der Umstand, dass das Produkt aus der Praxis und von PraktikerInnen entwickelt und betreut wird.

Einige Elemente der Kritik basieren auf noch unvollständiger Kenntnis bzw. mangelndem Verständnis des Konzeptes von Klassencockpit. *Üben, Klassencockpit als Hausaufgabe, Entwickeln von Schattennotensystemen, Deutung schlechter Ergebnisse bloss als Ergebnis einseitiger, scheinbar nicht lernzielkonformer Aufgaben, Verwendung von Klassencockpit-Noten fürs Zeugnis und Zwang der Schulbehörde zum Einsatz von Klassencockpit usw.* sind Beispiele von inadäquatem Einsatz des Instrumentes. Damit eine Subkultur der falschen Handhabung sich nicht (weiter) ausbreitet, müssen das Wesen und die Einsatzarten von Klassencockpit auch künftig zeitlich, medial und in der ästhetischen Form stimmig kommuniziert werden. Änderungen und Verbesserungen an den Modulen sollen nach dem jeweiligen Erkenntnisstand vorgenommen werden.

Testinhalte und Bewertung als vorrangige Elemente der Kritik

Inhaltliche Ausrichtung der Testfragen und *Notengebung* stehen konkret am häufigsten im Zentrum der Kritik und von Änderungswünschen. Es kann davon ausgegangen werden, dass ein Teil der Kritik sich verflüchtigen wird – dannzumal, wenn noch mehr Lehrpersonen über eine längere Praxis verfügen und erkennen, dass Klassenscockpit vor allem *über längere Zeit genutzt* den maximalen Ertrag erbringt und wenn Zweckbestimmung und Grenzen des Instrumentes allen klar sind.

In den Tests angesprochene Inhalte sind immer sehr sorgfältig in Bezug auf ihre Angemessenheit (d.h. mit Blick auf die ökonomische und soziale Heterogenität der St. Galler Gemeinden) zu prüfen. Da Testunterschiede ohnehin auch vom sozialen Kontext der SchülerInnen generiert werden, sollten in den Augen einiger Beobachter, wenigstens soweit dies vermeidbar ist, nicht unnötig Bereiche abgefragt werden, die sozial bedingte Leistungsunterschiede betonen. Allzu spezielle thematische Felder sollten demnach wohl vermieden werden oder wenigstens nur einen Teil der Tests ausmachen.

Der offenbar kaum vermeidbaren, durch einige Lehrpersonen vorgenommenen Vermischung von individueller Notengebungspraxis und Klassenscockpit-Note sollte entgegengesteuert werden, indem der Unterschied zwischen den beiden noch besser erklärt wird: Die Klassenscockpit-Note schafft den Bezug zur Leistung einer grösseren Gruppe von SchülerInnen aus dem ganzen Kanton, die individuelle Notengebung ist eine stärker der Verzerrung unterliegende, oft auf die Leistung der jeweiligen Klasse Bezug nehmende Bewertung einer individuellen Leistung.

Potenzial

Klassenscockpit beinhaltet das Potenzial, die Mentalität der an Lehrmitteln hängenden Lehrerinnen und Lehrer zu beeinflussen und sie zu motivieren, vermehrt auch selber lernzielbasierte, auf Fertigkeiten ausgerichtete Tests zu konzipieren. Viele Lehrpersonen haben realisiert, dass die Testaufgaben *anders* („*originell*“) sind, und sie suchen selber ebenfalls Aufgaben dieses Typs zu formulieren, der oft den kombinierten Einsatz von Wissen und Fertigkeiten erfordert. Klassenscockpit hat auch hier innovatives Potenzial, indem implizit zur Reflexion der Beurteilungspraxis aufgefordert wird. So wird vielleicht der Diskurs darüber intensiviert, ob es sinnvoll ist, eine ansatzweise gelöste Aufgabe mit einem Viertelpunkt zu bewerten, oder dass auch der Zugang zur Mathematik letztlich über sprachliche Fertigkeiten erschlossen werden muss, und so weiter.

Klassenscockpit, so möglicherweise das wichtigste Fazit der externen Evaluation, befindet sich an der Schnittstelle der laufenden Bemühungen, mit standardisierten Vergleichsmessungen aus einer Tendenz der Beschönigung von mittelmässigen Leistungen herauszufinden und in Richtung einer verbesserten Bildungsgerechtigkeit zu gehen. Die gewählte Strategie, den Lehrpersonen die notwendigen Lernprozesse in einem geschützten Rahmen zu ermöglichen, scheint einer der wichtigen Erfolgsfaktoren von Klassenscockpit zu sein. Wenn künftig noch mehr Module bereit stehen, die noch mehr Fertigkeiten des Curriculums testen, wird auch der Vorbehalt der Einseitigkeit zunehmend an Boden verlieren.

Der korrekten und bewussten Nutzung der Möglichkeiten von Klassencockpit stehen gewisse Unarten gegenüber, die den Nutzen von Klassencockpit beeinträchtigen. Hier braucht es noch einiges an Aufklärungs- und Informationsarbeit, was mittelfristig einige der noch skeptischen Lehrpersonen zu Interessierten machen könnte. Offene Instrumente haben aber immer die Schwäche des Missbrauchs. Es wird nicht zuletzt auch an den Nutzerinnen und Nutzern selber, an der Schulpraxis (inkl. Schulleitung und Behörde) liegen, welcher Gewinn für Schule, Unterricht und Qualität von Lernprozessen durch Klassencockpit abfällt.

5. Empfehlungen

Aufgrund der erhobenen Befragungsdaten und eigener Erwägungen zu den vorliegenden Resultaten empfehlen wir Folgendes:

1. Klassenscockpit ist als profiliertes Produkt für die Selbstevaluation des Unterrichts und allgemein im Rahmen der Bemühungen um Schulqualität und Bildungsgerechtigkeit zu positionieren.
2. Die Module sind so universell zu konzipieren, dass sie auch ausserhalb des Kantons St. Gallen angewendet werden können (Deutschschweiz, deutschsprachige Nachbarn).
3. Im Zuge des auch in anderen Kantonen erfolgenden Einsatzes von Klassenscockpit sollte die (je kantonale) Stichprobenvergrösserung zur Eichung von Modulen geprüft werden.
4. Die Stufenmodule sind zu überdenken und allenfalls durch ein weiteres Klassenmodul zu ersetzen.
5. Die Zahl der Module pro Klasse sollte mittelfristig leicht erhöht und der eingeschlossene Bereich von Lerninhalten erweitert werden.
6. Mittelfristig ist die Ausdehnung auf ausgewählte weitere Lernbereiche (z.B. Fremdsprachen) zu prüfen.
7. Die Instruktion der Lehrpersonen ist zu überdenken. Es sind Wege zu suchen, wie das Verständnis der Lehrpersonen von den Möglichkeiten und Grenzen des Klassenscockpit verbessert werden kann.
8. Dazu sind insbesondere die Zugangsarten zu Klassenscockpit (Schulblatt/Bestellung, Kollege/Kollegin, Internet, Auftrag durch Schulleitung usw.) zu rekonstruieren um besser abschätzen zu können, wo ein Zugang zum Produkt ohne vorgängige sorgfältig Information möglich ist.
9. Lehrpersonen (und Behörden) sind auch verstärkt über unzweckmässigen oder missbräuchlichen Einsätze des Klassenscockpit aufzuklären.