

Lehrereinschätzungen zu zentralen Tests und Leistungsrückmeldungen

Ein Vergleich zwischen Baden-Württemberg und Thüringen

Kurzvorstellung erster Ergebnisse auf der 7. Tagung „Empiriegestützte Schulentwicklung“ in Mainz, 6./7. Dez. 2007

Dr. Uwe Maier

Pädagogische Hochschule Schwäbisch Gmünd

Mitarbeiterinnen: Dipl.Päd. Sibylle Jäger, Michaela Drexler

Kontakt: uwe.maier@ph-gmuend.de

Projektförderung durch:

- DFG: Möglichkeiten und Grenzen der Nutzbarmachung von Bildungsstandards und zentralen Lernstandserhebungen für eine ergebnisorientierte Unterrichtsentwicklung an weiterführenden Schulen
- PostdoktorandInnenprogramm der Landesstiftung Baden-Württemberg: Bildungsentscheidungen im Kontext von Bildungsstandards und externen Schulleistungstests

Problemstellung

Bildungspolitische Reaktion auf PISA:

- Evaluationsbasierte Steuerungskonzepte: z.B. Bildungsstandards und zentrale Vergleichsarbeiten (KMK 2002)
- Unterschiedliche Test- und Rückmeldesysteme in den einzelnen Bundesländern (Möglichkeit von Feldexperimenten)

Funktionen zentraler Tests (v.a. Vergleichsarbeiten)

- Bildungsmonitoring (Schulsystemebene)
- Qualitätssicherung (Einzelschulebene / Lehrer)
- Individualdiagnostik (Schülerebene)

Problem: Doppelfunktion von Evaluation (Altrichter/Heinrich 2006)



Wie informativ und handlungsleitend sind zentrale Leistungsrückmeldungen für Lehrkräfte?

Theoretische Überlegungen

Bedingungen einer datenbasierten Schul- und Unterrichtsentwicklung

- Allgemeine Komponenten- und Prozessmodelle für eine datengeleitete Schulentwicklung (Visscher & Coe 2003; Helmke & Hosenfeld 2005); Tresch 2007)
- Institutionelle Rahmenbedingungen auf Einzelschulebene (O'Day 2004; Abelmann & Elmore 1999): internal accountability
- Lernpsychologische Feedbackmodelle (Kluger/DeNisi 1996; Hattie/Timperley 2007): Lenkt Rückmeldung die Aufmerksamkeit auf die Aufgabe oder auf das Selbstkonzept?
- Fachdidaktisch-konzeptionelle Überlegungen zu Rückmeldungen (Büchter/Leuders 2005; Blum et al. 2005; Peek et al. 2006; Lorenz 2005)
 - Kompetenzmodelle und kriteriale Rückmeldungen
 - Faire Vergleiche
 - Aufgaben zwischen psychometrischen und fachdidaktischen Anforderungen
 - Fehlertypenanalyse
 - Zeitpunkt des Tests und der Rückmeldung: Zeit für Interventionen

Zentrale Tests in Baden-Württemberg und Thüringen (Klasse 6)

	Daten- kontrolle und Auswertung	Ausführ- lichkeit der Rück- meldung	Bezugs- norm	Fairer Vergleich	Support durch zentrale Stellen
BW Vergleichs- arbeiten (LS Stuttgart)	Lehrer	niedrig	sozial	nein	niedrig
Thüringen Kompetenz- tests (Uni Jena)	zentral	hoch	sozial und kriterial	ja	hoch

Ziele und Forschungsfragen

(Gesamtprojekt)

1. Akzeptanz und Rezeption von zentralen Tests und Rückmeldeinformationen erfassen
2. Abschätzen, in welchen Bereichen des professionellen Lehrerhandelns die Rückmeldeinformationen genutzt werden können
3. Können bestimmte Merkmale des Test- und Rückmeldesystems für die Akzeptanz und Nutzung der Daten verantwortlich gemacht werden?
4. Voraussetzungen für die professionelle Nutzung zentraler Leistungsrückmeldungen auf individueller (Lehrer) und institutioneller Ebene (Einzelschule)

Methodische Vorgehensweise

Lehrerperspektive:

- Quantitativ: Fragebogen
- Qualitativ: Interviews

Veränderungen:

Quasi-längsschnittliche Befragungen in BW (2006, 2007)

Varianz durch Merkmale des Test- und Rückmeldesystems:

Quantitative Befragung 2007 wurde auf Thüringen ausgedehnt

Exemplarisches Ergebnis der Interviewstudie (BW 2006)

Frage im Lehrerinterview:

Auswirkungen der Testrückmeldung auf zukünftigen Unterricht?

Aussagekategorien für Mathematik (n=36 Lehrkräfte an HS und RS):

- | | |
|---|----|
| • Keine nennenswerten Auswirkungen: | 20 |
| • Bestimmte Aufgabenstellungen werden übernommen: | 9 |
| • Mehr Wiederholungen: | 7 |
| • Arbeitstechniken verstärkt üben: | 4 |

Stichprobe der quantitativen Lehrerbefragung

		Angeschriebene Schulen		Klassen (geschätzt)		Rücklauf Fragebögen	
		2006	2007	2006	2007	2006	2007
BaWü	HS	161 (12,8 %)	577 (45,9%)	240	860	113	260
	RS	53 (12,3 %)	212 (49,3%)	170	680	81	272
	GY	42 (11,1 %)	200 (53,1%)	160	740	113	293
ges. BW		256 (12,4 %)	989 (47,9%)	570	2280	307	825
Thü ²	RegSch		120 (51,9%)		360		207
	GY ³		56 (50,2%)		280		103
	ges. Thü.		167 (50,7%)		640		310

Fragebogen

Erhalten Sie durch die Ergebnisse der VA zusätzliche (!) Hinweise für Ihre zukünftige Unterrichtsgestaltung?

Die Vergleichsarbeit gab mir zusätzliche Hinweise, ...

	trifft zu				trifft nicht zu
... welche Inhalte in Zukunft verstärkt behandelt werden sollten.	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... welche Aufgabenstellungen in Zukunft besser geübt werden müssen.	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... welche Lerninhalte häufiger wiederholt werden müssen.	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... ob die Reihenfolge der behandelten Stoffgebiete geändert werden sollte.	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... welche neuen Aufgabenstellungen ich in meinen Unterricht einbinden muss.	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... welche Stoffgebiete neu aufgenommen werden sollten.	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... welche Lernmaterialien (z.B. Bücher, Arbeitshefte,...) für meinen zukünftigen Unterricht gut geeignet sind.	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Messung der Lehrereinschätzungen

- Allgemeine Akzeptanz zentraler Tests (6 Items, alpha = .89)
- Test als Belastung (4 Items, alpha = .79)
- Curriculare Validität (4 Items, alpha = .84)
- Diagnostische Hinweise (5 Items, alpha = .89)
- Hinweise für Notengebung (5 Items, alpha = .80)
- Hinweise für zukünftige Übungen/Wiederholungen (3 Items, alpha = .90)
- Hinweise auf inhaltliche Änderungen (4 Items, alpha = .77)

Ratingskalen von 1= Ablehnung bis 5 = Zustimmung

Messung des schulinternen Umgangs mit zentralen Testrückmeldungen

- Welche Gremien/Personengruppen haben Einblick in die Klassen- bzw. Schülerergebnisse?
- In welchen Gremien/von welchen Personengruppen wurde wie intensiv über die Rückmeldungen diskutiert?
- Welche Informationen wurden an die Schüler bzw. Eltern weitergereicht?

Ergebnisse: Vergleich Thüringen – BW Lehrereinschätzungen

	Land	n	M	SD	Signifikanz
Allgemeine Akzeptanz von Vergleichsarbeiten	BW	816	2,68	1,08	p < 0.001
	THÜR	306	3,47	0,90	
Vergleichsarbeiten als Belastung	BW	816	2,69	1,01	p < 0.001
	THÜR	306	2,06	0,87	
Lehrplanvalidität der Vergleichsarbeit	BW	816	2,85	0,94	p < 0.001
	THÜR	306	3,45	0,84	
Diagnostische Hinweise	BW	826	2,59	1,06	p < 0.001
	THÜR	309	3,19	0,94	
Hinweise Notengebung	BW	827	2,81	1,00	p < 0.001
	THÜR	308	2,62	0,85	
Hinweise zukünftige Wiederholungen und Übungen	BW	819	2,93	1,22	p < 0.001
	THÜR	307	3,60	0,97	
Hinweise auf inhaltliche Änderungen	BW	821	2,08	0,89	p < 0.001
	THÜR	307	2,32	0,90	

Ergebnisse: Vergleich Thüringen – BW Lehrereinschätzungen nach Fach

		Allg. Akzeptanz	Belastung	Curriculare Validität
BW	Deutsch	2,54	2,75	2,75
	Mathematik	2,86	2,62	2,94
THÜR	Deutsch	3,21	2,23	3,36
	Mathematik	3,65	2,00	3,58
Gesamt	Deutsch	2,67	2,65	2,87
	Mathematik	3,12	2,41	3,16

Zweifaktorielle Varianzanalysen:
Signifikante Haupteffekte Land und Fach; keine Interaktionseffekte

Ergebnisse: Vergleich Thüringen – BW Lehrereinschätzungen nach Fach

		Diagn. Hinweise	Hinweise Noten- gebung	Hinweise Wieder- holungen	Hinweise Inhalte
BW	Deutsch	2,49	2,76	2,78	1,98
	Mathematik	2,72	2,91	3,09	2,18
THÜR	Deutsch	3,01	2,47	3,58	2,30
	Mathematik	3,34	2,73	3,77	2,37
Gesamt	Deutsch	2,59	2,70	2,94	2,04
	Mathematik	2,93	2,85	3,32	2,25

Zweifaktorielle Varianzanalysen:

Signifikante Haupteffekte Land und Fach; keine Interaktionseffekte

Ergebnisse: Vergleich Thüringen – BW Orte und Qualität schulinterner Diskussionen

	BW		THÜR		Sign. der Differenz (t-Test) p
	Anteil	Rating	Anteil	Rating	
Ausgewählte Kollegen	76%	1,24	67%	1,23	0,864
Parallelkollegen	81%	1,38	78%	1,51	0,000
Klassenkonferenz	40%	1,17	51%	1,22	0,187
Jahrgangstufenkonferenz	14%	1,25	25%	1,17	0,185
Fachkonferenz	52%	1,43	84%	1,61	0,000
Gesamtlehrerkonferenz	22%	1,14	39%	1,27	0,004
Schulkonferenz	6%	1,10	15%	1,11	0,945
Schulleitung	62%	1,23	72%	1,34	0,002
Schüler	82%	1,40	90%	1,57	0,000
Eltern	50%	1,11	72%	1,20	0,002

Rating der Qualität der Gespräche:

- 1 ... informelles Gespräch
- 2 ... systematische Diskussion

Diskussion

- Deutliche Länderunterschiede in sämtlichen Dimensionen
- Kompetenztests in Thüringen werden eher akzeptiert und können für Lerndiagnosen verwendet werden
- Vergleichsarbeiten in BW liefern den Lehrkräften eher selektionsdiagnostische Hinweise
- Bessere Nutzung der Rückmeldungen in Mathematik als „testfreundlicherem“ Fach
- Deutlich mehr Lehrkräfte in Thüringen berichten von eher systematischen Diskussionen in Fach- und Gesamtlehrerkonferenzen

Problematik:

- Lediglich Prüfung eines „Bündels“ von Rückmeldemerkmalen
- Keine Daten auf Schulebene

Literatur

- Abelmann, Charles & Elmore, Richard, with Even, J., Kenyon, S., & Marshall, J. (1999). When accountability knocks, will anyone answer? (CPRE Research Report No. RR-42). Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania.
- Blum, Werner; Christina Driike-Noe, Dominik Leiß, Bernd Wiegand & Alexander Jordan (2005). Zur Rolle von Bildungsstandards für die Qualitätsentwicklung im Mathematikunterricht. Zentralblatt für Didaktik der Mathematik (ZDM), 37,(4), 267-274.
- Büchter, Andreas & Leuders, Timo (2005b). From students' achievement to the development of teaching: requirements for feedback in comparative tests. In: Zentralblatt für Didaktik der Mathematik (ZDM), 37(4), 324-334.
- Hattie, John & Helen Timperley (2007). The Power of Feedback. Review of Educational Research, Vol. 77, No. 1, 81-112.
- Helmke, A. & Hosenfeld, I. (2005). Standardbezogene Unterrichtsevaluation. In Gerold Brägger, Beat Bucher, Norberg Landwehr (Hrsg.), Schlüsselfragen zur externen Schulevaluation (S. 127-151). Bern: Hep Verlag.
- Kluger, A.N., & DeNisi, A. (1996). The effects of Feedback Interventions on performance: A historical review, a meta-analysis, and a preliminary Feedback Intervention Theory. Psychological Bulletin, 119/2, 254-284.
- Lorenz, J.H. (2005). Zentrale Lernstandsmessung in der Primarstufe: Vergleichsarbeiten Klasse 4 (VERA) in sieben Bundesländern (Central student assessment in primary schools: comparative tests for grade 4 in seven federal states of Germany). Zentralblatt für Didaktik der Mathematik (ZDM), 37,(4), 317-324.
- O'Day, Jennifer A. (2004). Complexity, Accountability, and School Improvement. In: Fuhrman, S.H. & Elmore, R.F. (2004) (Eds.). Redesigning Accountability Systems for Education. New York/London: Teachers College Press. S. 15-43.
- Peek, Rainer; Pallack, Andreas; Döbelstein, Peter; Fleischer, Jens & Leutner, Detlef (2006). Lernstandserhebungen 2004 in Nordrhein-Westfalen – zentrale Testergebnisse und Perspektiven für die Schul- und Unterrichtsentwicklung. In: F. Eder, A. Gastager und F. Hofmann (Hrsg.). Qualität durch Standards, Waxmann-Verlag, Münster; S. 219-233.
- Tresch, Sarah (2007). Potenzial Leistungstest. Wie Lehrerinnen und Lehrer Ergebnisrückmeldungen zur Sicherung und Steigerung ihrer Unterrichtsqualität nutzen. Bern: hep-Verlag.
- Visscher, Adrie J. & Coe, Robert (2003). School performance feedback systems: Conceptualisation, Analysis, and Reflection. In: School effectiveness and school improvement, Vol. 14(3), 321-349.