
Leistungsrückmeldungen im Längsschnitt – Erste Erfahrungen mit dem Schüler-Monitoring-System (SMS)

Wilfried Bos / Nils van Holt

Gliederung

1. Ausgangslage
2. Einbettung des SMS
3. Instrumente
4. Rückmeldungen
5. Befunde aus der Erprobungsphase
6. Potentiale und Grenzen



Ausgangslage

- Idee:
 - Lernstandserhebungen / Vergleichsarbeiten als Versuch evidenzbasierte Unterrichtsentwicklung zu befördern
- Problematik:
 - Anforderungsüberfrachtung dieser System
 - Besonders individualpädagogische Sicht der Lehrkräfte kann oft nur randständig bedient werden, wegen der mangelnden Genauigkeit der Testergebnisse auf Klassen- insbesondere aber auf Schülerebene
- Alternative: Schüler-Monitoring-System (SMS)?

Schulen im Team

- Netzwerkbasierendes Schulentwicklungsprojekt
- Projektträger: Stiftung Mercator, Projektdurchführung und wissenschaftliche Begleitung: Institut für Schulentwicklungsforschung (IFS) in Kooperation mit dem MSW des Landes NRW
- Förderzeitraum: 3,5 Jahre (02/2007 – 07/2010) + 1 Jahr Verlängerung
- 40 Schulen in Netzwerken à 3-5 Schulen (10 Netzwerke insgesamt)
- Arbeit an selbst gewählten Entwicklungsbedarfen
- Vorgabe: Fachbezogene Unterrichtsentwicklung
- fachliche & organisatorische Unterstützung durch das IFS

Einbettung des SMS

- Kooperation zwischen Cito und dem IFS (Schulen im Team)
- Zwei Netzwerke arbeiten im Rahmen der Netzwerkarbeit gemeinsam mit Cito an der Entwicklung des SMS
- 14 Projektschulen sind an der Testentwicklung des SMS für die 5. und 6. Jahrgangsstufe beteiligt
- Domänen: Mathematik und Leseverständnis
- Ziel: objektiviertes Instrument zur systematischen Kontrolle der Lernentwicklung auf Individualebene

Das SMS konkret

- Längsschnittlich auf 2 Jahre angelegt
 - formatives Instrument

- 4 MZP jeweils am Anfang und am Ende eines Schuljahres
 - Vom Beginn des 5. bis Ende des 6. Jahrgangs
 - Also zu Beginn eines neuen Bildungsabschnittes (Sekundarstufe)

- Rückmeldungen erfolgen auf Schul-, Klassen- und Individualebene

- Mathematiktest
 - Subdomänen: Zahlen und Operationen, Messen, Zeit und Geld, Proportionalitäten und Bruchzahlen
 - Durch Cito und die Beteiligten Lehrkräfte erstellt oder aus dem Niederländischen übernommen und durch die Lehrkräfte geprüft, um curriculare Validität sicherzustellen
 - Rotiertes Testdesign (Multimatrix)
 - Adäquate Testdauer für die Schüler, trotz insgesamt vieler Aufgaben

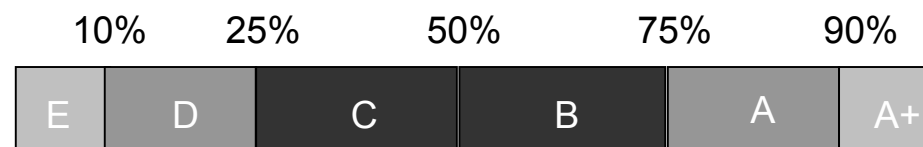
Instrumente

- Leseverständnistest
 - Domäne Lesekompetenz
 - Größtenteils Texte aus IGLU, die sich demnach am Literacy-Konzept orientieren
 - Multiple-Choice und offene Antworten
 - Literarische- und Sachtexte
 - Sachtexte überrepräsentiert
 - Auch hier aus testtheoretischen und zeitökonomischen Gründen Multimatrix-Design



Rückmeldungen

- In Anlehnung an Large-Scale-Assessments Transformation des Mittelwerts auf 500, da Lehrkräfte diese Darstellungsweise bereits kennen
- Nur Schüler die mindestens 50% der Aufgaben bearbeitet haben werden für Berichterstattung berücksichtigt
- Darstellung der Ergebnisse auf empirisch ermittelten Fähigkeitsniveaus



- Unterschieden nach zwei Normgruppen (Gym / GS), um Häufungen der Schulformen in bestimmten Niveaus zu vermeiden

Rückmeldungen

Rückmeldungen auf...

- Schulebene:
 - Vergleich an der Gesamtstichprobe anhand der Fähigkeitsniveaus und Normgruppen
- Klassenebene:
 - Punktwerte, Konfidenzintervalle, Prozent korrekt, Verteilung auf die Fähigkeitsniveaus
- Individualebene:
 - Individuelle Leistungsentwicklungen über bis zu vier MZP (im Rahmen von SiT 2 MZP)

Problem der Individualrückmeldungen

- Ergebnisse auf Individualebene zu berichten ist äußerst voraussetzungsvoll
 - Hinreichende Anzahl von Testaufgaben durch Multimatrixdesign
 - Trotzdem erhebliche Messfehler (vgl. Bos & Voss 2008)
 - Zusätzliche Absicherung durch längsschnittliche Betrachtung
 - Messungenaugigkeiten können so leichter identifiziert werden

Befunde aus der Erprobungsphase

1. Ergebnisse zur Qualität des Instrument
2. Explorative Ergebnisse zu Erfahrungen mit den Ergebnisrückmeldungen

quantitativ: Befragung zu den
Leistungsrückmeldungen (Koordinatoren und Fachlehrer)
qualitativ: 6 Experteninterviews zum Umgang mit den
Rückmeldungen



Skalierung

- Skalierung erfolgte anhand von zwei unterschiedlichen Verfahren parallel
 - Cito: One-Parameter-Logistic-Model (OPLM)
 - Verbindet die Vorteile des Raschmodells mit der Möglichkeit für die Itemparameter unterschiedliche Trennschärfeparameter zu bestimmen
 - Schulen im Team: Raschmodell
 - Setzt voraus: Itemhomogenität, lokale stochastische Unabhängigkeit, gleiche Trennschärfeparameter

Technische Details

- Abgleich der Verfahren: Rangfolgenkorrelation der Itemschwierigkeitsparameter nach Spearman-Rho
 - Koeffizienten:
 - Mathematik: 0.946 (Abweichung: 1,25 Rangplätze)
 - Leseverständnis: 0.906 (Abweichung: 1,64 Rangplätze)

- Testreliabilitäten:
 - Mathematik: 0.906
 - Leseverständnis: 0.742

Technische Details

- Gute Passung zwischen Aufgabenschwierigkeiten und Personenfähigkeiten
- Probleme: Differenzierung im oberen Fähigkeitsbereich (insbesondere Mathematiktest)
 - Beim zweiten MZP wurden die Tests um Aufgaben im oberen Fähigkeitsbereich ergänzt
- Aber: Insgesamt entsprechen die Tests den Standards internationaler Schulleistungsforschung



Erste Explorationen zu den Rückmeldungen

- Befragung fand im Rahmen einer Rückmeldeveranstaltung im Anschluss an die 1. Erhebungswelle statt
- Maximal zwei Netzwerkkoordinatoren pro Schule – die nachher auch als Multiplikatoren fungierten – konnten an der Veranstaltung teilnehmen
 - Daher ist der Rücklauf, obwohl von allen beteiligten Schulen Daten vorhanden sind, lediglich N=17
- Themen der Befragung:
 1. Verständlichkeit der Rückmeldungen
 2. Bedeutung für die eigene Arbeit
(Nützlichkeit, Konsequenzen, Kooperation)
- Ergänzt werden die Angaben der Koordinatoren durch Ergebnisse der Fachlehrer als Hauptadressaten der Rückmeldungen (N=37)

Bewertung der Verständlichkeit

	MW	SD
Sind die Ergebnismeldungen für Sie verständlich und übersichtlich dargestellt?	3,3	0,47
Finden sie die vorgelegten Beispielgrafiken verständlich?	3,4	0,51
War die schriftliche Erklärung zu den Ergebnismeldungen verständlich?	3,2	0,56

Antwortskalierung (1 = *gar nicht verständlich* bis 4 = *sehr verständlich*)

Bedeutung für die eigene Arbeit

	MW	SD
Bedeutung der Ergebnismeldung für die unterrichtspraktische Arbeit	3,0	0,66
Nützlichkeit der Rückmeldungen hinsichtlich der Diagnose des Lernstandes der Schülerinnen und Schüler	3,4	0,61
Nützlichkeit hinsichtlich der Verbesserung und Weiterentwicklung des eigenen Unterrichts	2,9	0,66

Antwortskalierung (1 = trifft nicht zu bis 4 = trifft zu)

Bedeutung für die eigene Arbeit

	MW	SD
Rückmeldungen werden als Anlass zum Austausch mit einzelnen Kollegen genommen	3,4	0,61
Rückmeldungen werden zur Diskussion in der Fachkonferenz genutzt	3,5	0,72
Es besteht eine generelle Bereitschaft Konsequenzen für den Unterricht aus den Rückmeldungen zu ziehen	3,5	0,72
Aus den Rückmeldungen lassen sich konkrete Maßnahmen ableiten	2,9	0,80
Die Rückmeldungen decken Stärken und Schwächen der Schülerinnen und Schüler auf	3,0	0,93
Es wird durch die Rückmeldungen besser als zuvor deutlich wie einzelne Schülerinnen und Schüler zu fördern sind	2,6	0,99

Antwortskalierung (1 = *trifft nicht zu* bis 4 = *trifft zu*)



Ergänzende Befunde

- Die Ergebnisse der Fachlehrerbefragung stützen die der Koordinatorenbefragung für den Bereich Mathematik
 - 73% geben an, dass die Rückmeldungen in der Fachgruppe diskutiert werden
 - 65% der Lehrkräfte sehen einen hohen Nutzen für den eigenen Unterricht
- Die Aussagen aus den Interviews stützen diese Ergebnisse ebenfalls
 - Zusätzlich wird die Finanzierung des SMS problematisiert
 - Eignung zur Diagnostik und individuellen Förderung wird gesehen
 - z.T. für diese Perspektive für gut befunden, z.T. aber auch zu wenig detailliert (nur 3 Subdomänen)



Zusammenfassung

- Insgesamt deuten die Ergebnisse auf ein Potential des SMS hin
 - Insbesondere, da die Befragung im Anschluss an den 1. MZP stattfand und die Aussagekraft des SMS im Zuge weiterer MZP erheblich zunimmt, da auch Entwicklungen berichtet werden können
- Weitere Vorteile des SMS:
 - Daten werden extern erhoben und ausgewertet
 - Erhebung erfolgt längsschnittlich (4 MZP)
 - Individualrückmeldungen sind präzise durch Multimatrix-Design
 - Keine verpflichtende Teilnahme
 - Keine Verpflichtung zur Rechenschaftslegung

Potenziale und Grenzen: Evaluation

- Das SMS bietet das Potential eines formativen Evaluationsinstruments, da es längsschnittlich am Anfang eines neuen Bildungsabschnittes eingesetzt wird und die Rückmeldungen relativ kurz nach der Datenerhebung erfolgen
 - Aber:
 - Aufgabenformate bieten nur begrenzte Diagnosemöglichkeiten durch vorwiegend geschlossene Antwortformate

Potenziale und Grenzen: Koordinierung

Das SMS ist ein marktförmiges Instrument, Angebot und Nachfrage entscheiden also über die Einführung (kann im Zuge sinkender Schülerzahlen das Portfolio der Schule stärken)

Aber:

Gemeinschaft

Gemeinschaftliche
Entscheidung für die
Einführung
Kooperative Weiterentwicklung
des Unterrichts

vs.

Hierarchie

- Aufktoyierung durch z.B. die Schulleitung
- Gefahr: Lehrkräfte sehen das SMS als Beurteilung der eigenen Arbeit
- Nicht intendierter Effekt: Widerstand



Potentiale und Grenzen: Implementierung

IFS

Grenzen:

- Anknüpfung an die Bildungsstandards besteht bisher nicht
- Instrumente sind noch in der Entwicklung
- Die Kosten müssen von den Schulen getragen werden
- Bisher keine kriterialen und fairen Vergleiche möglich
- Begleitung und Qualifizierung der Beteiligten erforderlich

Entwicklungsprozess:

- Anpassung der Test an den oberen Leistungsbereich ist erfolgt
- Signifikante Fort- und Rückschritte der einzelnen Schüler werden zusätzlich berichtet
- Einführung von Kompetenzstufen wird vorbereitet
- Seminare zur Implementierung werden angeboten

- Erforschung der Wirksamkeit des SMS ist dringend erforderlich
 Durch: quantitative und qualitative Befragungen, Beobachtung der Verarbeitungsprozesse, Ausweitung der Forschung auf alle beteiligten Akteure, Prüfung der Leistungsentwicklung
- Systematische Begleitung des Rückmelde- und Rezeptionsprozesses durch externe Berater
 Stadtweise Einführung des SMS in Kooperation mit den RBB's, dadurch kontinuierliche Beratung und Qualifizierung

Vielen Dank für die Aufmerksamkeit!

Kontakt: officebos@ifs.tu-dortmund.de
 vanHolt@ifs.tu-dortmund.de

- Bos, W./Voss, A. (2008): Empirische Schulentwicklung auf Grundlage von Lernstandserhebungen – Ein Plädoyer für einen reflektierten Umgang mit Ergebnissen aus Leistungstests. In: Die Deutsche Schule 100, H.4, S. 449 - 458.
- Berkemeyer, N., van Holt, N., (in Druck). Leistungsrückmeldungen im Längsschnitt – Erste Erfahrungen mit dem Schüler-Monitoring-System (SMS). Wacker, A., Maier, U., Wissinger, J. (Hrsg.). *Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung – Empirische Befunde und forschungsmethodische Implikationen*. VS Verlag.