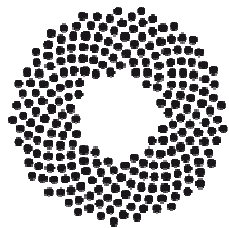


18. EMSE in Soest

Glanz und Elend einer Strategie des Bildungsmonitorings

Norbert Maritzen



ifbq

Institut für Bildungsmonitoring
und Qualitätsentwicklung



1. Bildungsmonitoring als gesamtstaatliche Aufgabe der KMK – Glanz und Elend in historischer Perspektive
2. Linien einer Reformulierung der Strategie zum Bildungsmonitoring
3. Bildungsmonitoring und Forschungsfragen

1. Bildungsmonitoring als gesamtstaatliche Aufgabe der KMK – Glanz und Elend in historischer Perspektive
2. Linien einer Reformulierung der Strategie zum Bildungsmonitoring
3. Bildungsmonitoring und Forschungsfragen

Konstanz und Plön: Revolution von der Peripherie aus...

Konstanzer Beschluss (1997): Leistungsvergleiche innerhalb Deutschlands in der SEK I mit Blick auf grundlegende Kompetenzen

Plöner Beschlüsse (2006): Gesamtstrategie zum Bildungsmonitoring

1. Teilnahme an internationalen Schulleistungs-untersuchungen (PIRLS/IGLU, TIMSS, PISA)
2. Überprüfung des Erreichens der Bildungsstandards im Ländervergleich (IQB-Länder-vergleich)
3. Vergleichsarbeiten zur landesweiten Überprüfung der Leistungsfähigkeit einzelner Schulen (VERA 3 und VERA 8)
4. gemeinsame nationale Bildungsberichterstattung von Bund und Ländern

... und gut 15 Jahre später?

- Gemischte Bilanz mit Leistungen und Brüchen
- Potenzial für die Rückgewinnung von Steuerungs-fähigkeit, aber auch Restriktionen der Wahrnehmung gesamtstaatlicher Verantwortung durch die Ländergemeinschaft
- Probleme der Finanzierbarkeit in Zeiten der Haushaltskonsolidierung
- „Unterbestimmtheit“ der Nutzung der Ergebnisse auf allen Ebenen
- Ungeklärter Status des Wissens, das mit Large-Scale-Assessments produziert wird, im politischen Diskurs

- Rückbezug auf einen normativen, empirischer Überprüfung zugänglichen Referenzrahmen (an empirische Forschung anschlussfähiger Bildungsbegriff, Bildungsstandards, Kompetenzmodelle)
- Minimalkonsens über ein länderübergreifendes Transparenzprogramm
- tendenzielle Delegitimierung von reinen „Meinungen“ durch Evidenzansprüche
- Veränderung der Aufmerksamkeitsstrukturen von Öffentlichkeit und Medien
- Agenda-Setting im politischen und öffentlichen Raum:
 - a. Bildung, Humankapital und internationaler Wettbewerb, b. Bedeutung von Bildung im Lebenslauf; c. Bildungsgerechtigkeit; d. Risikogruppen und deren Kosten für den Sozialstaat und eine demokratische Gesellschaft; e. Zuwanderung, Integration und soziale Kosten; f. regionale Disparitäten in Deutschland und Bildungsgerechtigkeit; g. Föderalismus als Verantwortung und/oder Wettbewerbsgemeinschaft;

- Einstieg aller Länder in die Schulleistungsevaluation auf Basis standardisierter Tests (VERA):
 - a. länderübergreifende Kooperation auf fachlicher Ebene, b. Institutionalisierung der Testadministration, c. Professionalisierung des Personals, d. Weiterentwicklung der Rückmeldeformate, e. Rückwirkung auf Referenzmaßstäbe von Lehrkräften (Einübung des Umgangs mit kriterialen Bezugsnormen)
- Stärkung der Erklärungskraft der internationalen Assessments durch strategische Erweiterungen
- infrastrukturelle Absicherung der Large-Scale-Assessment durch Gründung des IQB und des ZIB
- Definition von konsensfähigen, zukunfts-offenen Handlungsfeldern nach Veröffentlichung der ersten PISA-Ergebnisse
- systematische Verstärkung und institutioneller Ausbau der empirischen Bildungsforschung und deutliche Steigerung ihrer internationalen Visibilität

- Variation theoretischer Modelle insbesondere zur Modellierung sozialer Herkunftsmerkmale
- widersprüchliche Befunde in Trendanalysen
- fehlendes Equating zwischen internationalen Assessments und den IQB-Ländervergleichen
- Reduktion der Entwicklung von Kompetenzstufenmodellen für die IQB-Länder-vergleiche (Hauptschule? Schreiben? integrierte HSA/MSA-Modelle als „Notlösung“ ...)
- Umgang mit der zunehmenden Anzahl von Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf (Beteiligung, Testkonstruktion, Berichterstattung)
- Heterogenität der Durchführungsbedingungen von VERA → Risiko der Diskreditierung und Korrumpierung

- VERA auf der Basis von Kompetenzstufenmodellen → Risiko einer unzulässigen Verknüpfung der VERA-Befunde mit denen des IQB-Ländervergleichs
- Kalibrierung und Normierung der VERA-Tests unter Bedingungen, die im Anschluss an den flächendeckenden Einsatz der Tests schwer händelbare Varianzen in den Befunden produzieren (unplausible Varianz über die Zeit, zwischen den Domänen bzw. Subdomänen etc.)
- Risiko des Legitimationsverlustes: Wie ist bei gleichem Referenzrahmen z. B. zu erklären, dass der Anteil der Schülerinnen und Schüler, die keinen Hauptschulabschluss erhalten, weit unter dem in Assessments festgestellten Anteil der Schülerinnen und Schüler liegt, die die Mindeststandards nicht erreichen?

Strukturelle Grenzen der Verständigungsbereitschaft und Koordinationsfähigkeit der Länder

- Übersichtskataloge von Maßnahmen, die die Länder ohnehin durchführen, wenn finanzielle Aufwendungen nötig, die nicht zu länderspezifischen Prioritätensetzungen passen
- Zustimmung zu länderübergreifendem Handlungsprogramm (z. B. BISS), sofern die Länder den Status quo ihrer ohnehin finanzierten Maßnahmen fortschreiben können und der Bund die zusätzlichen Kosten komplett übernimmt
- Tabuisierung von Schulstrukturfragen, da entsprechende länderspezifische politische Festlegungen kaum folgenreich problematisierbar sind

- Tabuisierung von Maßnahmen, die mit hohem Risiko politischer „Abstrafung“ verbunden sind (z. B. Akademisierung des Personals in der frühkindlichen Erziehung, der Struktur der Lehrerbesehung und der Lehrerarbeitszeit, der Aufwertung des Grundschullehreramt, der Veränderung insgesamt der Personalstruktur an Schulen)
- „Pflege“ eines unvollendeten Verhältnisses zum Bildungsmonitoring
- keine Lösung für persistente Problemlagen: Leseförderung als Querschnittsaufgabe, Rückstand der Jugendlichen mit Migrationshintergrund, geschlechtsspezifische Leistungsunterschiede, geringe Besetzung der Leistungsspitze...)

Bildung als intergenerationelle, gesamtgesellschaftliche Aufgabe stößt im Kontext bildungsföderaler staatlicher Verfasstheit auf deutliche Grenzen der Bearbeitbarkeit.

1. Bildungsmonitoring als gesamtstaatliche Aufgabe der KMK – Glanz und Elend in historischer Perspektive
2. Linien einer Reformulierung der Strategie zum Bildungsmonitoring
3. Bildungsmonitoring und Forschungsfragen

- Zeitpunkt der Überprüfung im Primarbereich (Jg. 4)
- Klärung, ob **a.** getrennte Überprüfung der HSA-Standards in Jg. 8, **b.** integrierte Kompetenzstufenmodelle in Jg. 9 oder **c.** nur MSA-Kompetenzstufenmodelle in Jg. 9 und keine Überprüfung der HSA-Standards
- Klärung der internationalen Anbindung des IQB- Ländervergleichs (Stichprobe statt Ankeritems)
- Klärung, ob die Anbindung der Ländervergleiche an die internationalen LSA fachlich sinnvoll ist
- Ergänzung um die 2012 beschlossenen Bildungsstandards für die Sek. II
- Zielsetzung und Durchführung von VERA aktualisieren um den Beschluss der KMK zur Weiterentwicklung von VERA (2012); Unterscheidung von Ankoppelung und Anbindung aufheben; Text um weitere Ergänzungen aktualisieren z.B. Aufgabenentwicklung durch das IQB etc.
- div. redaktionelle Aktualisierungen
- ...

- Beratungsprozess: Wie organisiert die KMK den Strategieentwicklungsprozess? als Event-Hinterstübchen-Walzer, als kooperativer Prozess zwischen Wissenschaft, Administration und Politik, als...?
- Handlungsrahmen: Welche Erfahrungen wurden mit dem Spannungsverhältnis zwischen politischem Auftrag und wissenschaftlicher Berichterstattung gemacht? Welche Freiheitsgrade soll künftig ein Bildungsmonitoring haben?
- normativer Referenzrahmen: Welchen Überarbeitungsbedarf ergibt sich konzeptionell und angesichts empirischer Befunde bei den Bildungsstandards?
- Nutzungsvoraussetzungen: Welches theoretische Modell haben wir davon, wie VERA auf die Unterrichts- und Schulentwicklung wirkt? Gibt es für postulierte Wirkzusammenhänge empirische Evidenz? Wie kann die Entwicklungsfunktion von VERA gestärkt werden? noch konsequentere Abkopplung vom Monitoring, längsschnittliche Anlage, adaptives Testen, Verknüpfung mit Prozess- und Kontextdaten...

Linien einer Reformulierung (1)

- breit angelegter Diskussionsprozess, auch unter Nutzung der in den Instituten akkumulierten Expertise
- Entscheidung zwischen „Rolle vorwärts“ und „Rolle rückwärts“: Bekenntnis zu den Steuerungsimplicationen des Bildungsmonitorings („evidence based policy“ nicht nur als Beschwörungsformel)
- systematische Auswertung der Erfahrungen mit den Bildungsstandards → realistisches Konzept von Leistungserwartungen
- von der Ranking-Philosophie zur Fokussierung auf gemeinsam zu bewältigende Problemlagen → Veränderung der Berichterstattung, Konzept für Ergänzungsstudien
- institutionelle Absicherung und Schaffung nachhaltiger Finanzierungsstrukturen für das IQB

- Beratungsprozess: Wie organisiert die KMK den Strategieentwicklungsprozess? als Event-Hinterstübchen-Walzer, als kooperativer Prozess zwischen Wissenschaft, Administration und Politik, als...?
- Handlungsrahmen: Welche Erfahrungen wurden mit dem Spannungsverhältnis zwischen politischem Auftrag und wissenschaftlicher Berichterstattung gemacht? Welche Freiheitsgrade soll künftig ein Bildungsmonitoring haben?
- normativer Referenzrahmen: Welchen Überarbeitungsbedarf ergibt sich konzeptionell und angesichts empirischer Befunde bei den Bildungsstandards?
- Nutzungsvoraussetzungen: Welches theoretische Modell haben wir davon, wie VERA auf die Unterrichts- und Schulentwicklung wirkt? Gibt es für postulierte Wirkzusammenhänge empirische Evidenz?

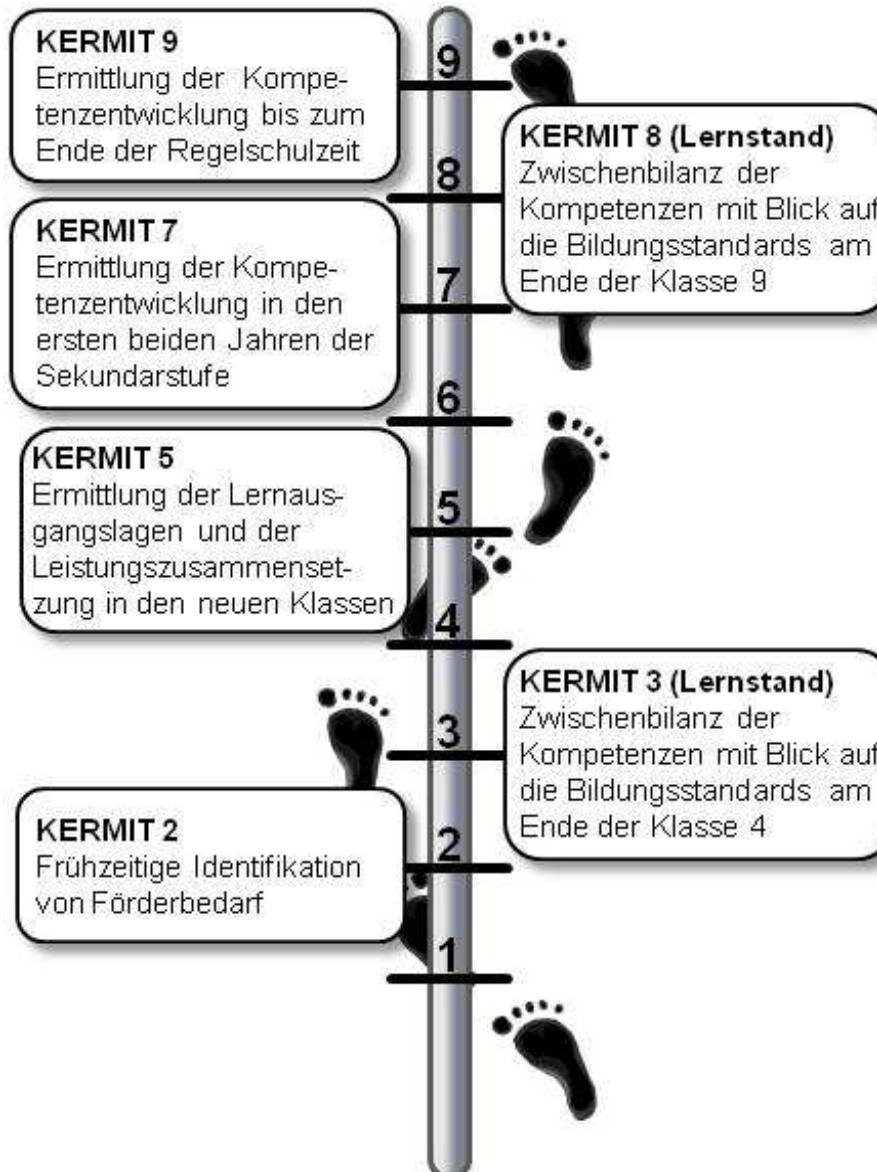
Linien einer Reformulierung (2)

- breit angelegter Diskussionsprozess, auch unter Nutzung der in den Instituten akkumulierten Expertise
- Entscheidung zwischen „Rolle vorwärts“ und „Rolle rückwärts“: Bekenntnis zu den Steuerungsimplicationen des Bildungsmonitorings („evidence based policy“ nicht nur Beschwörungsformel)
- systematische Auswertung der Erfahrungen mit den Bildungsstandards → realistisches Konzept von Leistungserwartungen
- von der Ranking-Philosophie zur Fokussierung auf gemeinsam zu bewältigende Problemlagen → Veränderung der Berichterstattung, Konzept für Ergänzungsstudien
- institutionelle Absicherung und Schaffung nachhaltiger Finanzierungsstrukturen für das IQB
- **systematische Stärkung der Entwicklungsfunktion von VERA → Konsolidierung der Infrastruktur, noch konsequentere Abkopplung der Testformate von denen des Systemmonitorings, längsschnittliche Anlage, adaptives Testen, Verknüpfung mit Prozess- und Kontextdaten, Verknüpfung mit schulinterner, formativer Evaluation...**

KERMIT – Kompetenzen ERMITteln

Referenzrahmen Hamburger Bildungspläne

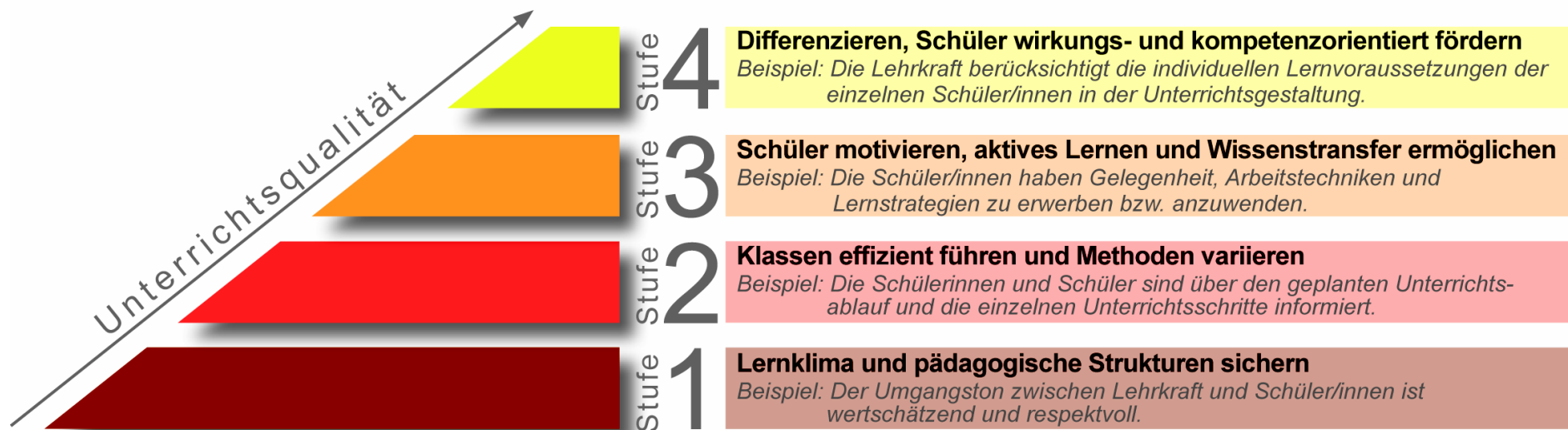
- Tests in De, En, Ma, NW
- alle Schüler
- externe Testleiter
- jährlich
- echter Längsschnitt
- Rückmeldung von Lernständen und -entwicklungen



Referenzrahmen Bildungsstandards

Gelegenheitsstruktur: Unterricht

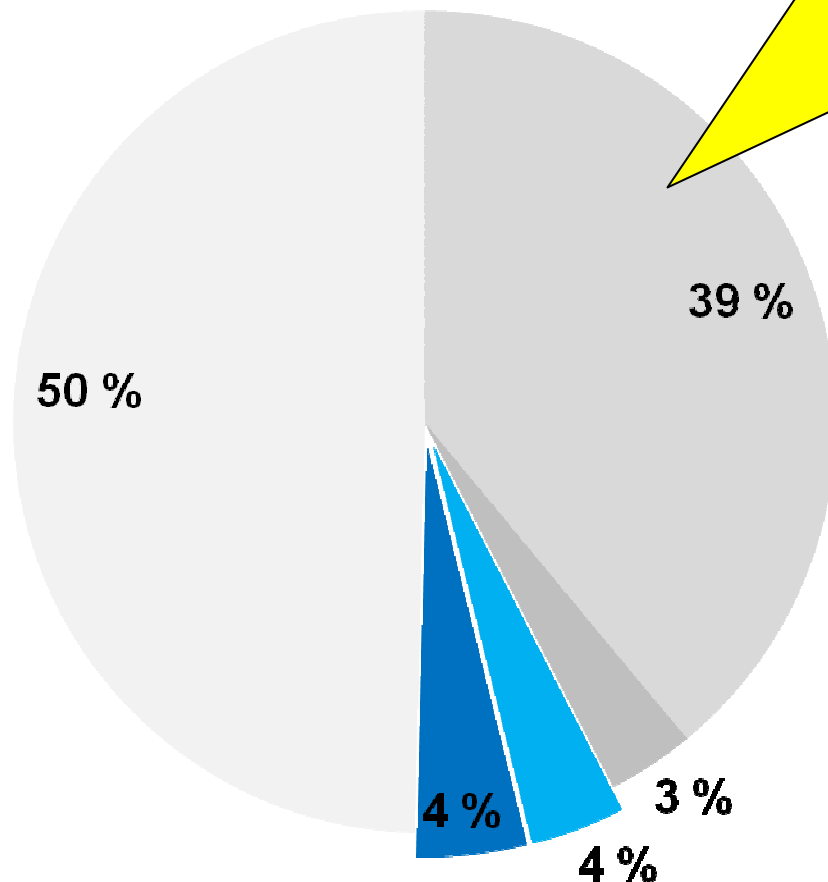
- Grundlage zz. ca. 27.000 beobachtete Unterrichtseinheiten
- Repräsentative Stichproben auf System- und Schulebene
- Standardisierte Unterrichtsbeobachtungen mit geschulten Beobachtern (60-110 U-Einheiten je Schule à 30 Items)
- Probabilistisches, kriterial bestimmtes Stufenmodell (Rasch-Skalierung)



Zusammenhang von Unterricht und mittleren Abiturleistungen

Datengrundlage

- Unterrichtsbeobachtungen der Schulinspektion in Sekundarschulen mit Oberstufe (Sj 2010/11)
- Abiturergebnisse in Deutsch, Mathematik und 1. Fremdsprache der Sekundarschulen auf Schulebene 2011



Unterschiede in den Abiturergebnissen erklären sich

- zu 39 % über die soziale Zusammensetzung der Schülerschaft,
- zu 3 % über Unterschiede in den Schwierigkeiten der Abiturfächer,
- **zu 8 % über Unterschiede in der beobachteten Unterrichtsqualität** und
- zu 50 % aus anderen (nicht beobachteten) Merkmalen.

Fremdsprache)

■ **Qualität der beobachtbaren Unterrichtsmerkmale**

■ **Streuung der beobachtbaren Unterrichtsmerkmale**

■ **andere Merkmale (z. B. Tiefenstruktur des Unterrichts)**

Praktische Bedeutung der Unterschiede in der Unterrichtsqualität

- Auch an Schulen mit schwächeren Leistungsergebnisse kann es „guten“ Unterricht geben.
- An Schulen mit guten Leistungsergebnisse findet man eher selten schwachen Unterricht, da an diesen Schulen anscheinend Vorkehrungen getroffen wurden, die es ermöglichen, großflächige Schwächen im Unterricht zu vermeiden.
- Es gibt systematische Zusammenhänge zwischen Unterrichtsqualität und Leistungsergebnissen: Mit je zehn Prozentpunkten mehr Unterricht der Stufe 3 geht eine Steigerung der mittleren Schülerleistung im Abitur um 0,25 Punkte einher.

Praktische Bedeutung der Unterschiede in der Unterrichtsqualität



Modellrechnung:

- Lägen beispielsweise an einer Schule A 100 Prozent der Unterrichtssequenzen auf Stufe 1 oder 2 und an einer Schule B 100 Prozent der Unterrichtssequenzen auf Stufe 3, würde die mittlere Abiturleistung der Schülerinnen und Schüler an Schule B wahrscheinlich allein aufgrund des Unterrichts um 2,5 Punkte höher ausfallen als an Schule A.
- Würde es allen weiterführenden Hamburger Schulen gelingen, einen überdurchschnittlich guten, effektiven Unterricht zu etablieren und Qualitätsunterschiede innerhalb der Schule zu reduzieren, würden wahrscheinlich knapp 1.000 der jährlich etwa 8.000 Hamburger Abiturientinnen und Abiturienten bessere Abiturergebnisse erzielen können.

1. Bildungsmonitoring als gesamtstaatliche Aufgabe der KMK – Glanz und Elend in historischer Perspektive
2. Fallbeispiel Hamburg – Standardsicherung als integriertes System der Datengewinnung und des Daten-Feedbacks
3. Bildungsmonitoring und Forschungsfragen

Woher rührt der Ruf nach „konkreteren Antworten“ aus der Forschung?

- unrealistische Erwartungen der Politik – uneingelöste Versprechen der Forschung
- eingeschränkte Verarbeitungskapazität von Schulen und Schuladministration für Monitoring-Daten; quantitatives/qualitatives Problem
- Hattie-mania
- Fehlvorstellungen von Möglichkeiten / Grenzen, Handlungslogiken und Zeithorizonten empirischer Bildungsforschung
- Unkenntnis über gängige Verfahren kompetitiver Forschungsförderung
- Fehlen von Vermittlungsinstanzen zwischen Wissenschaft und Praxis
- Überdruß an der Deskription von Problemlagen → Was wissen wir, wenn wir das wissen? Erklärungsabstand zwischen „PISA“ und Handlungserfordernissen „vor Ort“
- Unklarheit über den Status des Wissens, das im Bildungsmonitoring generiert wird

Wissensformen (Prenzel et al. 2009)

Beschreibungswissen informiert unter anderem über vorherrschende Bedingungen des Unterrichts (von Klassengrößen bis Materialien), über vorfindbare Unterrichtsmuster und über Unterrichtsergebnisse (Kompetenzen, Interessen);

Vorhersagewissen beschreibt Beziehungen zwischen Prädiktoren (z.B. Klassengröße, Klassenzusammensetzung, Vorwissen) und Kriteriumsvariablen (z.B. Leistung, Belastungserleben der Lehrkraft), wobei die Beziehungen aufgrund der beobachteten korrelativen Zusammenhänge nicht als kausale Beziehungen interpretiert werden können;

Erklärungswissen identifiziert Faktoren (z.B. Rückmeldungen der Lehrkraft, Unterrichtsmuster), die als kausal relevant gelten für die Entwicklung bestimmter Merkmale (z.B. Lernmotivation, Kompetenzentwicklung);

Veränderungswissen gibt an, wie vorgegangen werden kann, um unter gegebenen Umständen (z.B. große Leistungsunterschiede in einer Klasse) mit bestimmten Mitteln (z.B. individualisierter Unterricht) bestimmte Ziele (z.B. Bildungsstandards) zu erreichen;

Zielwissen ist Wissen, das Beziehungen zwischen über- und untergeordneten Zielen (Kompetenzentwicklungsmodelle) oder mögliche Zielkonflikte (z.B. Leistungsorientierung versus soziale Orientierung) betrifft;

Bewertungswissen stellt Verbindungen her zwischen Ziel- und Beschreibungswissen und resultiert aus der Evaluation von Lernumgebungen.

Wissensformen (Prenzel et al. 2009)

Beschreibungswissen informiert unter anderem über vorherrschende Bedingungen des Unterrichts (von Klassengrößen bis Materialien), über vorfindbare Unterrichtsmuster und über Unterrichtsergebnisse (Kompetenzen, Interessen, etc.);

Large-Scale-Assessments

Vorhersagewissen beschreibt Beziehungen zwischen Unterrichtsbedingungen (z.B. Klassenzusammensetzung, Vorwissen) und Kriteriumsvariablen (z.B. Leistung, Belastungserleben der Lehrkraft), wobei die Beziehungen aufgrund der beobachteten korrelativen Zusammenhänge nicht als kausale Beziehungen interpretiert werden können;

Erklärungswissen identifiziert Faktoren (z.B. Rückmeldungen der Lehrkraft, Unterrichtsmuster), die als kausal relevant gelten für die Entwicklung bestimmter Merkmale (z.B. Lernaktivitäten, Kompetenzentwicklung);

**Ergänzungsstudien /
längsschnittliche
Kontrollgruppendesigns**

Veränderungswissen gibt an, wie vorgegangene Unterrichtsmaßnahmen (z.B. große Leistungsunterschiede in einer Klasse) bestimmte Ziele (z.B. Bildungsstandards) zu erreichen;

Zielwissen ist Wissen, das Beziehungen zwischen über- und untergeordneten Zielen (Kompetenzentwicklungsmodelle) oder möglichen Lernaktivitäten (z.B. soziale Orientierung) betrifft;

**Implementationsstudien,
Begleitforschung**

Bewertungswissen stellt Verbindungen her zwischen Ziel- und Beschreibungswissen und resultiert aus der Evaluation von Lernumgebungen.

Mit der Kenntnis von Ursachen steigt der politische Druck, Problemlagen spezifisch und adäquat anzugehen. Ist das wirklich gewollt?

Vielen Dank

Norbert.Maritzen@ifbq.hamburg.de
www.hamburg.de/bsb/ifbq/

