

Timo Leuders, Freiburg

Von der Klassenarbeit bis zur Leistungsstudie – Chancen und Risiken

1. Empiriegestützte Unterrichtsentwicklung: Der neue Weg?

In den letzten Jahren blicken die deutschsprachigen Bildungssysteme zunehmend auf die Fachleistungen von Schülerinnen und Schülern. Den internationalen *large scale assessments* TIMSS und PISA folgten regionale wie etwa LAU oder MARKUS. Hinzu treten seit einigen Jahren zentrale Leistungstests wie die Lernstandserhebungen in NRW (Klasse 9) oder die länderübergreifende VERA (Klasse 4), die empirisch reliable Rückmeldungen für Schülerleistungen im Lerngruppenvergleich generieren. Schließlich ist ebenfalls eine Verstärkung zentraler Prüfungen auszumachen.

Diese vielfältigen Formen von Leistungsmessung haben erhebliche Rückwirkungen auf das Schulsystem. Im Sinne einer qualitätssichernden Funktion werden positive Rückwirkungen auf Unterricht und auf Schülerleistungen erwartet. Aber auch kritische Stimmen melden sich zu Wort und weisen auf die Gefahren hin, wie etwa die Einengung des Bildungsgedankens auf das fachliche und messbare, die unerwünschten Nebenwirkungen eines „teaching to the test“, die unzureichende Qualität der Rückmeldungen und den Mangel an konkreter Unterstützung von Lehrerinnen und Lehrern bei ihrer Professionalisierung (Bonsen/Gathen 2004, MNU 2004, 2006).

Die Einschätzung wie Chancen und Risiken zu gewichten sind, wie die einen genutzt und die anderen vermieden werden können, gehen unter den Protagonisten der Bildungssysteme weit auseinander. Im Rahmen eines Minisymposiums haben sich Vertreter verschiedener deutschsprachiger Bildungssysteme (Deutschland, Schweiz, Luxemburg) unter verschiedenen Perspektiven (Standardsetzung, zentrale Leistungsmessung, Abschlussprüfungen, diagnostische Tests) zur Frage nach Chancen und Risiken der neueren Entwicklungen in der Leistungsmessung ausgetauscht.

2. Leistungsmessung: Ein weites Feld!

Das Thema Leistungsmessung in Schulen (Weinert 2001) umspannt ein weites Spektrum an Instrumenten und Funktionen. Die Gegenüberstellung in der Tabelle weiter unten deutet bereits an, dass es vielfältiger Überlegungen bedarf, welches Instrument welcher Funktion dienen kann: Überschreiten zentrale Leistungsstudien ihren Gültigkeitsbereich, wenn aus ihnen individualdiagnostische Informationen abgeleitet werden? Können Abschlussprüfungen messtheoretischen Kriterien unterworfen werden, wenn

die Lernbedingungen von Schülerinnen und Schülern nicht vergleichbar sind? Können Portfolios, Präsentationsprüfungen usf. Leistungsdimensionen erfassen, die in Tests ausgeblendet bleiben? Können Leistungsskalen bildungsrelevante Kompetenzen hinreichend modellieren?

| Instrumente der Leistungsmessung | Funktionen der Leistungsmessung |
|---|---|
| <ul style="list-style-type: none"> ▪ Klassenarbeiten ▪ Vergleichsarbeiten ▪ Portfolios, Facharbeiten & Co. ▪ Zentrale Abschlussprüfungen ▪ Zentrale Vergleichsarbeiten ▪ Leistungsstudien | <ul style="list-style-type: none"> ▪ Individualdiagnose ▪ Lerngruppendiagnose ▪ Vergabe von Abschlüssen ▪ Lerngruppenvergleich ▪ Bildungsmonitoring ▪ Kompetenzmodellierung |

Die Diskussion dieser Zusammenhänge und die Entwicklung geeigneter Instrumente ist künftig gemeinsame Aufgabe von Psychometrie, Bildungsforschung, Schulentwicklungsforschung und Fachdidaktik. Insbesondere gilt es, bei jeder einzelnen Maßnahme über eine angemessene Gewichtung der Kriterien zu diskutieren:

- Messtheoretische Dignität, z.B. in Form von hinreichender Reliabilität, bedeutsamen Effektstärken, Modelltransparenz
- Inhaltliche Validität z.B. hinsichtlich der didaktischen Tragfähigkeit der Kompetenzmodelle oder der Wiedergabe des Bildungsgedankens
- Nützlichkeit für Unterrichtsentwicklungsprozesse, Brauchbarkeit und Verständlichkeit für die tägliche Arbeit von Lehrkräften
- Wirksamkeit im Rahmen der Entwicklung von Schulen und Bildungssystemen

Zu den drängenden Fragen gehören dabei u.a.: Welche Veränderungen hinsichtlich Leistungsmessung kommen auf Lehrkräfte zu? Wie werden die Instrumente nach Funktion und Struktur ausgeschärft? Welche Kompetenzmodelle haben/brauchen wir für welchen Zweck? Wie wird empirische Leistungsmessung in Fachentwicklung umgesetzt?

Literatur

- Bonsen, M./ Gathen, J. (2004): Schulentwicklung und Testdaten. Die innerschulische Verarbeitung von Leistungsrückmeldungen. In: Jahrbuch der Schulentwicklung, (2004) 13, S. 225-252
- Büchter, A./ Leuders, T./ Bruder, R. (2005). Quality development in mathematics education by focussing on the outcome: new answers or new questions? Zentralblatt für Didaktik der Mathematik, 4/2005 (August), S.263-266.
- MNU (2004, 2006). Empfehlungen zur Umsetzung des Bildungsstandards im Fach Mathematik / ... zur Evaluation von Schülerkompetenzen im Fach Mathematik auf der Grundlage der Bildungsstandards der KMK. MNU 8/04 bzw. 6/06 (Einhefter)
- Weinert, F.-E. (2001). Leistungsmessungen in Schulen. Weinheim und Basel: Beltz.