Forschungsmonitor Schule
Rezensionen für die Praxis

Sonja Hensel

Rezension zu

Kippers, W. B., Wolterinck, C. H. D., Schildkamp, K., Poortman, C. L. & Visscher, A. J. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and Teacher Education, 75*, 199–213.

Kommentierter Kurzbefund

Die Autorengruppe um Wilma B. Kippers unterscheidet zwei Ansätze des formativen Assessments: Strategien des Assessment for Learning (AfL), um den laufenden Lernprozess zu verbessern, z. B. durch Fragen stellen und Klassenraumgespräche, sowie Data-Based Decision Making (DBDM), um lernförderliche Innovationen zu implementieren und zu evaluieren, z. B. auf Basis standardisierter Tests und Umfragen. Doch wie ist die Praxis? Hierzu wird folgenden Fragen nachgegangen:

1. Welche Instrumente und Prozesse des formativen Assessments werden am häufigsten in der Unterrichtspraxis verwendet?
2. In welchem Maße werden Strategien des Assessments for Learning (AfL) und Data-Based Decision Making (DBDM) in der Unterrichtspraxis eingesetzt?
3. Welche Voraussetzungen halten Lehrkräfte für wichtig, um AfL und DBDM in der Unterrichtspraxis zu implementieren?

Die Fragen 1 und 3 wurden auf Basis von Interviews mit 12 Lehrkräften von 4 niederländischen Sekundarschulen untersucht, zur Beantwortung von Frage 2 wurden zusätzlich 479 Lehrkräftefragebögen von 27 Sekundarschulen analysiert.

Im Ergebnis nutzen alle 12 Lehrkräfte schriftliche Tests und stellen Fragen. Die Mehrheit verwendet Klassenraumgespräche (11), Hausaufgaben (10), Beobachtungen (9), Reflexionsstunden (9), Präsentationen (8), mündliche Tests (8) und Portfolios (7). Die Hälfte setzt Fragebögen (6) und praktische Aufgaben (6) ein, wenige digitale Tests (3).

Die häufigste AfL-Strategie besteht darin, Lernende über Lernziele und Erfolgskriterien zu informieren (M = 3.3, Skala: 1–5), jedoch geben in den Interviews nur 2 Lehrkräfte an, dies explizit zu tun. Fragen stellen und Klassenraumgespräche werden ebenfalls relativ häufig genutzt (M = 3.1), seltener wird Feedback gegeben (M = 2.8). Peer- und Selbsteinschätzungen werden kaum veranlasst (M = 1.8). Im Hinblick auf DBDM ziehen die Lehrkräfte systematisch erhobene Daten nicht sehr häufig zur Weiterentwicklung des Unterrichts heran (M = 2.9).

Für die erfolgreiche Implementierung von AfL und DBDM nennen die Lehrkräfte eine positive Haltung, detaillierte Rückmeldungen, Unterstützung durch die Schulleitung und Übereinstimmung mit dem Curriculum als wichtigste Voraussetzungen.

Die Studie bietet einen interessanten Einblick in die Nutzung von AfL und DBDM in der Unterrichtspraxis und verdeutlicht, dass in beiden Bereichen Entwicklungspotenziale bestehen, um Daten systematisch für Unterrichtsentwicklung zu nutzen.

CC BY NC ND

Hintergrund

Ein Konzept, auf das sich die Autorengruppe um Wilma B. Kippers bezieht, ist das des „Formative Assessment“, also der Einsatz von Instrumenten und Prozessen zur Datenerhebung und -nutzung, um das Lernen der Schülerinnen und Schüler zu unterstützen. Dabei unterscheiden sie zwei Ansätze:

1. Assessment for Learning (AfL), das auf täglichen Rückmeldungen aus Gesprächen, Beobachtungen etc. beruht und zu zeitnahen Reaktionen und Anpassungen in Bezug auf eine verbesserte Unterrichtspraxis führt, sowie
2. Data-Based Decision Making (DBDM), das systematisch schulische Datenquellen analysiert, darauf aufbauend Maßnahmen konzipiert, durchführt und evaluiert.

Beide Ansätze sind als iterative Prozesse zu denken, bei denen durch Erhebungen wie standardisierte Schulleistungstests oder Hausaufgaben Daten generiert werden, die von Lehrkräften und Lernenden als Feedback zu Lernprozessen und Lernbedingungen genutzt und in eigenes Handeln übersetzt werden können. Sie verfolgen das gemeinsame Ziel, Unterricht und Schule durch das Sammeln von Informationen besser an die Bedürfnisse der Schülerinnen und Schüler anzupassen.

Sie unterscheiden sich in der Häufigkeit und in der Systematik, mit der Daten erhoben und ausgewertet werden. Während AfL häufig und eher unsystematisch in der täglichen Arbeit von Lehrkräften stattfindet, beispielweise in Form von Unterrichtsgesprächen, Tests oder Hausaufgabenergebnissen, beruht DBDM auf systematisch erhobenen, hochwertigen Daten, die in einem zirkulären Prozess aus Zielbestimmung, Sammeln, Analysieren und Interpretieren der Daten und Umsetzung der Ergebnisse in Handeln verwendet werden.

Unklar ist, inwiefern die beiden Ansätze in der Praxis realisiert und ggf. kombiniert werden. Daher gehend die Forschenden folgenden Fragen nach:

1. Welche Assessment-Instrumente und -Prozesse werden am häufigsten in der Unterrichtspraxis verwendet?
2. In welchem Maße werden Strategien des Assessments for Learning (AfL) und Data-Based Decision Making (DBDM) in der Unterrichtspraxis eingesetzt?
3. Welche Voraussetzungen halten Lehrkräfte für wichtig, um AfL und DBDM in der Unterrichtspraxis zu implementieren?

Design

Die Forschenden nutzen bereits vorhandene Daten aus zwei Projekten, bei denen es darum ging, Lehrkräfte zu einer verstärkten Datennutzung anzuregen. Beide wurden in niederländischen Sekundarschulen durchgeführt (Lernende im Alter zwischen 12 und 18 Jahren) und bedienten sich verschiedener Erhebungsinstrumente, die zu einem Mixed- Method-Design in der vorliegenden Studie führten.

Zum einen wurden 479 Fragebögen zum Einsatz von AfL und DBDM ausgewertet, die von Lehrkräften von 27 Schulen ausgefüllt worden waren, die aus einem großen niederländischen Schulbezirk stammen. Das entspricht 52 % der für die Untersuchung eingeladenen Schulen. Männliche und weibliche

Lehrkräfte sind in der Stichprobe annähernd genauso verteilt wie in der Grundgesamtheit (49,5 % zu 49,7 % gegenüber 47,5 % zu 52,5 %). Bei anderen Merkmalen wie Schulgröße oder Trägerschaft der Schule gibt es teils erhebliche Unterschiede, indem zum Beispiel sehr große Schulen und Schulen in nicht-staatlicher Trägerschaft in der Stichprobe überrepräsentiert sind.

Der Fragebogen enthielt 42 Items, bei denen Antworten auf einer 5-stufigen Likert-Skala von *fast nie* (*seltener als in 10 % der Unterrichtsstunden*) bis zu *fast immer* (*in mehr als 90 % der Unterrichtsstunden*) angekreuzt werden konnten. Die Faktoren- und Reliabilitätsanalysen wurden mit SPSS durchgeführt. 10 Items wurden aus der Auswertung herausgenommen, weil sie eine Ladung < .5 aufwiesen. Vier weitere Items, die eine Ladung < .5 aufwiesen, wurden aus theoretischen Gründen beibehalten. Die Reliabilität der Skala *Lernabsichten und Erfolgskriterien kommunizieren* wird als ausreichend (c_a = .76) eingeschätzt, die der anderen 4 Skalen (*Daten für den Unterricht nutzen*, *Fragen stellen und Klassenraumgespräche*, *Feedback*, *Peer- und Selbsteinschätzung*) als gut (c_a > .80).

Für jede Skala wurden Mittelwerte und Standardabweichungen errechnet sowie eine Varianzanalyse (ANOVA) durchgeführt, um Unterschiede zwischen den Lehrkräften verschiedener Fächergruppen zu untersuchen.

Zum anderen wurden aus dem zweiten Projekt vier Schulen ausgewählt, an denen Interviews mit insgesamt 12 Lehrkräften (für die Hauptfächer Niederländisch, Englisch und Mathematik) geführt wurden und die außerdem Checklisten zum Einsatz von formativem Assessment ausfüllten. Die Schulen wurden unter anderem danach ausgewählt, ob sie überdurchschnittliche Ergebnisse in Lernstandserhebungen zeigten, weil die Forschenden davon ausgingen, dass in solchen Schulen AfL und DBDM genutzt würden und dementsprechend dort untersucht werden könne, wie diese genutzt werden. Es wurden sieben weibliche und fünf männliche Lehrkräfte mit einer Unterrichtserfahrung von weniger als 5 bis hin zu mehr als 25 Jahren befragt.

Die Interviews wurden mit Hilfe eines Leitfadens auf Niederländisch geführt, der 20 offene Fragen umfasste und zum Beispiel die Fragen danach enthielt, welche Formen der Leistungsmessung die Lehrkräfte in ihrer täglichen Unterrichtspraxis benutzen und wie sie den Schülern Feedback geben. Die Interviews wurden transkribiert und mit einem a priori erstellten Kodierschema mit Hilfe des Programms Atlas.ti von zwei Kodierern ausgewertet (Inter-Rater-Reliabilität: Cohen‘s Kappa = .69).

Für eine Triangulation der Daten füllten die interviewten Lehrkräfte außerdem zwei Checklisten aus, auf denen sie in einer Liste ankreuzten, welche Formen des Assessments sie in ihrem Unterricht einsetzen und welche (maximal fünf) Bedingungen sie für die Implementation von AfL und DBDM als wichtig erachten. Beide Checklisten waren zuvor von den Forschenden aufgrund des vorhandenen Forschungsstands erstellt worden.

Ergebnisse

Welche Assessment-Instrumente und -Prozesse werden am häufigsten in der Unterrichtspraxis verwendet?

Zur Beantwortung der ersten Frage werteten die Forschenden die Checklisten aus, die die 12 für die Interviews ausgewählten Lehrkräfte ausgefüllt hatten. Dabei zeigt sich, dass alle Lehrkräfte schriftliche Tests und das Stellen von Fragen als Assessments nutzen und ein großer Teil (8 und mehr) darüber hinaus Klassenraumgespräche, Hausaufgaben, die Beobachtung der Lernenden, Reflexionsstunden, mündliche

Tests und Präsentationen. Ebenfalls von den Forschenden zur Auswahl gegebene Möglichkeiten wie das Portfolio, Fragebögen, praktische und digitale Aufgaben wurden von maximal 7 Lehrkräften genutzt. Die Checkliste wurde von den Lehrkräften um Diskussionen im Klassenraum und Hörverstehensübungen ergänzt, die von zwei Lehrkräften bzw. einer Lehrkraft eingesetzt wurden.

In welchem Maße werden Strategien des Assessments for Learning (AfL) und Data-Based Decision Making (DBDM) in der Unterrichtspraxis eingesetzt?

Um die zweite Forschungsfrage zu beantworten, nutzten die Forschenden sowohl die Fragebögen als auch die Interviews als Datenquelle. Bei der Auswertung der Fragebögen zeigt sich, dass die Fragebogenskala *Lernabsichten und Erfolgskriterien kommunizieren* von den Lehrkräften am häufigsten gewählt wurde, so dass sich ein Mittelwert von $M = 3.30$ ergibt ($N = 419$, $SD = 0.79$), was einem Einsatz in 50 bis 75 % der Unterrichtstunden entspricht. Demgegenüber gaben in den Interviews nur 2 der 12 Lehrkräfte an, die Lernenden explizit über die Lernziele und Erfolgskriterien zu informieren. Als Begründung dafür, dies nicht zu tun, wurde geäußert, dass es dazu keine Notwendigkeit gebe oder diese Information einschüchternd wirken könne.

Einen Mittelwert von 3.07 ($N = 436$, $SD = 0.80$) erreicht die Skala *Fragen stellen und Klassenraumgespräche*, was wiederum einem Einsatz in 50 bis 75 % der Unterrichtsstunden entspricht. Hier zeigt sich ein signifikanter Einfluss von Fächergruppen: Lehrkräfte der Gesellschaftswissenschaften nutzen diese Art der Rückmeldung signifikant häufiger ($N = 83$; $M = 3.43$; $SD = 0.79$) als die der Geisteswissenschaften (Literaturwissenschaften, Religion, Philosophie) ($N = 132$; $M = 3.04$; $SD = 0.72$) und diese wiederum häufiger als die Lehrkräfte in den Naturwissenschaften ($N = 134$; $M = 2.96$; $SD = 0.78$) und als Lehrkräfte, die andere Fächer unterrichten ($N = 85$; $M = 2.92$; $SD = 0.85$). Wie schon beschrieben, wurde diese Form des Assessments auch in den Interviews häufig genannt.

Deutlich seltener angewendet wird von den Lehrkräften schriftliches Feedback ($N = 424$; $M = 2.82$; $SD = 0.86$), wobei hier unterschiedliche Formen von „Gut gemacht!“ bis zu differenzierten Rückmeldungen subsumiert werden. Der Mittelwert entspricht einer Verwendung in 25 bis 50 % der Unterrichtstunden. Lehrerinnen verwenden diese Form der Rückmeldung etwas häufiger als ihre männlichen Kollegen (leicht signifikantes Ergebnis). Auch bei den Antworten in den Interviews zeigt sich ein eingeschränkter Einsatz von Feedback – abgesehen von kurzen mündlichen Rückmeldungen an die Schülerinnen und Schüler während der Unterrichtstunden. Hier sehen die Forschenden Entwicklungspotential, insbesondere wenn es darum geht, Leistungen der Lernenden als Feedback für die eigene Unterrichtsqualität zu nutzen. Dies zu tun, gaben nur 5 Lehrkräfte in den Interviews an.

Am seltensten kreuzten die Lehrkräfte bei den Skalen des AfL im Fragebogen an, Methoden der Peer- und Selbsteinschätzung zu nutzen ($N = 396$; $M = 1.77$; $SD = 0.71$), was einem Einsatz in 10 bis 25 % der Unterrichtsstunden entspricht. In den Interviews nannten Lehrkräfte als Grund für die seltene Nutzung dieser Methoden mangelnde Fachkenntnisse oder mangelnde Reflexionsfähigkeit der Lernenden.

Im Bereich des DBDM (*Daten für den Unterricht nutzen*) ergibt sich, dass die Lehrkräfte eher selten (25 bis 50 % der Stunden) systematisch erhobene Daten für die Weiterentwicklung des eigenen Unterrichts nutzen ($N = 439$; $M = 2.85$; $SD = 0.75$). Die Auswertung der Interviews zeigt, dass die Lehrkräfte nur in beschränktem Maße die Daten nutzen, um in ihrem Unterricht Veränderungen vorzunehmen. Sie berichten von mündlichem oder schriftlichem (kurzen) Feedback bei einer negativen Leistungsentwicklung (bisweilen auch nur in Form einer Note) oder einer Anpassung des Unterrichts, um schwachen Lernenden besser gerecht zu werden.

Welche Voraussetzungen halten Lehrkräfte für wichtig, um AfL und DBDM in der Unterrichtspraxis zu implementieren?
In Bezug auf die dritte Forschungsfrage nannten die Lehrkräfte vier Voraussetzungen besonders häufig, die notwendig seien, um AfL und DBDM erfolgreich zu implementieren: eine positive Haltung der Lehrkräfte, eine detaillierte Rückmeldung aus dem Assessment (nicht nur Korrektheit der Lösung, sondern Aufschluss über den Lösungsweg), Unterstützung durch die Schulleitung und eine Übereinstimmung zwischen Assessment und Curriculum. Von einer Minderheit der Befragten wurden außerdem Wissen und Fähigkeiten in Bezug auf Unterricht, Zusammenarbeit der Lehrkräfte, die Integration des Assessments in den Fachunterricht, valide Instrumente und der Zugang zu Technologie genannt.

Diskussion und Einschätzung

Zum Hintergrund

Die Autorengruppe um Wilma B. Kippers verortet ihre Untersuchung zur schulischen Praxis von datengestützter Unterrichtsentwicklung im Forschungsbereich zu formativem Assessment. Dies birgt einerseits den Vorteil, etablierte Systematiken zugrunde legen zu können, um standardisierte Erhebungsinstrumente zu entwickeln. Insbesondere gilt dies für die zentralen Strategien des Assessments for Learning (AfL) und für die Kategorien der Voraussetzungen einer erfolgreichen Implementierung. Andererseits wird das Verständnis von formativem Assessment ungewohnt weit ausgelegt, indem Data-Based Decision Making (DBDM) subsumiert wird. Die Überdehnung wird u. a. daraus ersichtlich, dass Kippers et al. ihr DBDM-Konstrukt auf Unterricht beschränken (d*ata use for instruction*), obwohl DBDM ein allgemeiner Ansatz für Schulentwicklungsprozesse ist, die nicht zwangsläufig auf Unterricht und laufende Lernprozesse bezogen sein müssen.

Zum Design

Das Untersuchungsdesign wird sehr systematisch und ausführlich dargestellt, so dass die Konstruktion der Erhebungsinstrumente, die Auswahl und Zusammensetzung der Stichprobe und das Vorgehen bei der Auswertung gut nachvollziehbar sind.

Wenig aussagekräftig erscheint dagegen die quantitative Auswertung der Ergebnisse der Interviewstudie, vor allem da die Anzahl der Befragten mit 12 sehr klein ist. Ob ein Bewertungsinstrument von den Interviewten häufig oder selten genutzt wird, ist nicht generalisierbar, zumal gezielt Lehrkräfte ausgewählt worden waren, die bereits Erfahrungen mit datengestütztem, formativem Feedback hatten. Hier wäre eine ausführlichere Darstellung der Erklärungen und Begründungen der Befragten wünschenswert gewesen, weil diese einen Mehrwert gegenüber den Daten der Fragebogenstudie liefern.

Zu den Ergebnissen

Die Studie zeigt, dass vor allem schriftliche Tests und das Stellen von Fragen, oft auch Klassenraumgespräche, Hausaufgaben und Beobachtungen für Assessments genutzt werden. Seltener verwendet werden Portfolios, Fragebögen und digitale Methoden.

Was die Häufigkeit des Einsatzes von AfL-Strategien unter den Lehrkräften angeht, steht die Information über Lernabsichten und Erfolgskriterien an erster Stelle, aber selten werden diese explizit und ausführlich kommuniziert. Das Stellen von Fragen und Klassenraumgespräche werden ebenfalls recht häufig genutzt, wobei sich Unterschiede zwischen den Fächergruppen zeigen. Schriftliches Feedback kommt seltener zum Einsatz; hier zeigen sich jedoch Unterschiede zwischen den Geschlechtern. Peer- und

Selbsteinschätzungen werden selten verwendet, oft wegen von den Lehrkräften vermuteten mangelnden Fachkenntnissen und Selbstkompetenzen der Schülerinnen und Schüler. Systematisch erhobene DBDM-Daten werden selten zur Weiterentwicklung des Unterrichts genutzt.

Für die erfolgreiche Implementierung von AfL und DBDM sind eine positive Haltung der Lehrkräfte, die Erhebung detaillierter und aussagekräftiger Daten, Schulleitungsunterstützung und Übereinstimmung der Assessments mit dem Curriculum entscheidend.

Die Forschenden weisen abschließend darauf hin, dass AfL und DBDM gemäß den Ergebnissen nur einen geringen Stellenwert in der Unterrichtspraxis haben und deshalb ein verstärkter Fokus auf der Förderung solcher Praxis liegen sollte. Dazu seien Kriterien nötig, wie genau ein effektiver Einsatz formativen Feedbacks aussehe und wie Lernende dabei besser mit einbezogen werden können. Die in dieser Untersuchung gefundenen Unterschiede zwischen Lehrkräften verschiedener Fächergruppen und unterschiedlichen Geschlechts bieten ebenso Hinweise, wo Fortbildungs- und Implementationsbemühungen ansetzen können, wie die als bedeutsam identifizierten Rahmenbedingungen wie eine positive Einstellung der Lehrkräfte und die Unterstützung der Schulleitung.

In dieser Hinsicht liefert die Studie interessante Daten zum Ist-Stand von AfL und DBDM, die – auch, wenn sie aus den Niederlanden stammen – in ihrer Grundtendenz die Basis für Überlegungen zur Weiterentwicklung formativen Feedbacks in deutschen Schulen sein können.

Reflexionsfragen für die Praxis

Nachfolgende Reflexionsfragen sind ein Angebot, die Befunde der rezensierten Studie auf das eigene Handeln als Lehrkraft oder Schulleitungsmitglied zu beziehen und zu überlegen, inwiefern sich Anregungen für die eigene Handlungspraxis ergeben. Die Befunde der rezensierten Studien sind nicht immer generalisierbar, was z. B. in einer begrenzten Stichprobe begründet ist. Aber auch in diesen Fällen können die Ergebnisse interessante Hinweise liefern, um über die eigene pädagogische und schulentwicklerische Praxis zu reflektieren.

Reflexionsfragen für Lehrkräfte

- Wie beurteile ich den Stellenwert von AfL und DBDM für meinen Unterricht und für die Schulentwicklung?
- Welche Datenquellen nutze ich und welchen Stellenwert haben sie in meiner Unterrichtspraxis?
- In welchen Phasen der Unterrichtsplanung und Entwicklung kommen welche Arten formativen Feedbacks zum Einsatz?
- Wie nehme ich die Rahmenbedingungen an meiner Schule für den Einsatz von AfL und DBDM wahr?
- Was benötige ich, um AfL und DBDM zielgerichteter im Unterricht einzusetzen?

Reflexionsfragen für Schulleitungen

- Wie beurteile ich den Stellenwert von AfL und DBDM für die Unterrichts- und Schulentwicklung?
- Welche Datenquellen nutzen wir und welchen Stellenwert haben sie in unseren Entscheidungsprozessen?

- Wie beeinflusse ich die Rahmenbedingungen in meiner Schule für den Einsatz von AfL und DBDM?
- Wie motiviere ich die Lehrkräfte an meiner Schule, verstärkt verschiedene Formen von AfL und DBM einzusetzen?
- Inwieweit schaffe ich die Rahmenbedingungen, um langfristige Entwicklungsprozesse und Kooperationen zwischen den Lehrkräften zu ermöglichen?

Rezensent/-in

Dr. Sonja Hensel, Lehrerin am Berufskolleg in Siegburg sowie Lehrbeauftragte an der Universität Siegen. Arbeitsschwerpunkte: Rechtschreib-, Schreib- und Lesedidaktik, selbstreguliertes und kooperatives Lernen.

Zitiervorschlag

Hensel, S. (2025). Rezension zu Kippers, W. B., Wolterinck, C. H. D., Schildkamp, K., Poortman, C. L. & Visscher, A. J. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. Teaching and Teacher Education, 75, 199–213. *Forschungsmonitor Schule, 192*. Abgerufen von https://www.forschungsmonitor-schule.de/print.php?id=182

Urheberrecht

Dieser Text steht unter der CC BY-NC-ND 4.0 Lizenz. Der Name des Urhebers / der Urheberin soll bei einer Weiterverwendung wie folgt genannt werden: Sonja Hensel (2025) für den Forschungsmonitor Schule.