

Waldemar STRAUMBERGER, Bielefeld

Selbstdiagnosebögen als Grundlage für individuelles Üben

Selbstdiagnosebögen sind vom Prinzip ein Hilfsmittel zum eigenständigen Strukturieren von Übungsphasen. Durch die Darstellung der im Unterricht behandelten Kompetenzen ermöglichen sie es dem Lernenden, seinen Lernstand mittels eigener Einschätzung selbst zu bestimmen und seinen Übungsbedarf daraus abzuleiten. Die Übungsphase kann durch Übungsbedarf und die Übungsempfehlungen zu den Kompetenzen individuell strukturiert werden.

Der erste Bericht im deutschsprachigen Raum über die Nutzung von Selbstdiagnosebögen stammt von Rosel Reiff aus dem Jahr 2006, in dem sie über „Selbst- und Partnerdiagnose im Mathematikunterricht“ berichtet. Im Fokus stehen dabei ihre ersten Erfahrungen beim Einsatz in einer 7. Klasse einer Realschule. Kernpunkte sind dabei die Entwicklung der Einschätzung sowie die Motivation der Lernenden. Zu Beginn besteht bei Jungen eine Tendenz zur Überschätzung ihrer eigenen Leistung und bei Mädchen eine Tendenz zur Unterschätzung (Reiff 2006). Unabhängig vom Geschlecht wird die eigene Einschätzung mit der Zeit realistischer. Des Weiteren zeigen die Lernenden eine hohe Motivation beim Nutzen der Materialien. Ein weiterer Bericht aus dem Jahr 2011 stammt von Helmut Achilles, der Selbstdiagnosebögen zur Prüfungsvorbereitung in der Oberstufe einsetzte. Er berichtet ebenfalls von erhöhter Motivation seitens der Lernenden und betont zusätzlich den Nutzen solcher Materialien zur Förderung der Eigenverantwortung für das eigene Lernen (Achilles 2011).

Durch die Berichte stellt sich die Frage nach der Genauigkeit der Einschätzung. Gibt es zu Beginn Tendenzen zur Unter- oder Überschätzung und wie entwickelt sich diese? Gibt es Gruppen bei denen sich die Genauigkeit der Einschätzung unterscheidet?

Studie

Zum Untersuchen dieser Fragen wurde ein Studiendesign entwickelt, bei dem unter authentischen Bedingungen integriert in Übungsphasen Daten zur Selbsteinschätzung der Lernenden erhoben werden können. Die Übungsphasen haben einen Umfang von vier Unterrichtsstunden, die zur Vorbereitung auf Klassenarbeiten genutzt werden. Zu Beginn jeder Übungsphase schätzten sich die Lernenden selbst anhand der im Unterricht behandelten Inhalte ein und beginnen im Anschluss daran ihrem Übungsbedarf entsprechend zu Üben. In der dritten Stunde schätzen sich die Lernenden erneut ein und Bearbeiten einen Test zum Überprüfen ihres Lern-

In Institut für Mathematik und Informatik Heidelberg (Hrsg.), *Beiträge zum Mathematikunterricht 2016* (S. x–y). Münster: WTM-Verlag

standes. Für die Korrektur stand den Lernenden eine Musterlösung zur Verfügung, die die Aufgaben den Basiskompetenzen des Selbstdiagnosebogens zuordnete. Hiermit konnten die Lernenden ihre Testleistung mit ihrer Einschätzung vergleichen.

Die Stichprobe bestand aus der 5. Jahrgangsstufe einer Realschule in Bielefeld Mitte. In den drei Klassen der Jahrgangsstufe wurden die Inhalte parallel behandelt und auch die gleichen Klassenarbeiten geschrieben.

Im Folgenden werden die Daten zu vier Übungsphasen mit der oben beschriebenen Struktur dargestellt. Die Daten beziehen sich dabei auf 48 Lernende, bei denen zu den vier Übungsphasen die Daten zur Selbsteinschätzung vor dem Test und zur Testleistung vorlagen.

Methoden zur Auswertung

Für die Auswertung der Genauigkeit wurde auf Methoden der Forschung zu metakognitiven Urteilen zurückgegriffen. Bei der Forschung zu metakognitiven Urteilen liegt ein Fokus auf der Messung der Genauigkeit von Urteilen über die eigene Kognition. Bei der Messung werden zwei Formen von Genauigkeit unterschieden, die absolute Genauigkeit (absolute accuracy) und die relative Genauigkeit (relative accuracy). Bei der absoluten Genauigkeit steht die Passung zwischen Urteilen und Leistungen in Form von Differenzen im Vordergrund. Bei der relativen Genauigkeit hingegen wird das Verhältnis von Urteilen und Leistungen untersucht, z. B. durch Messung der Korrelation. Klassisch werden Urteile und Leistung überwiegend normativ mittels 0 für falsch und 1 für richtig codiert. Zum Teil wird die Einschätzung mit einer 10-stufigen Skala oder durch die Angabe eines Prozentwertes erhoben. Dementsprechend wurde die Einschätzung der vorliegenden Studie mittels 0; 0,33; 0,67 und 1 codiert. Zur besseren Vergleichbarkeit und aufgrund der beschränkten Anzahl der Items im Test zum Messen der Leistung wurde die Testleistung ebenfalls mittels 0; 0,33; 0,67 und 1 codiert. Dabei steht 0 für keine Aufgabe richtig bearbeitet, 0,33 für weniger als die Hälfte richtig bearbeitet, 0,67 für mindestens die Hälfte der Aufgaben richtig bearbeitet aber nicht alle und 1 für alle Aufgaben richtig bearbeitet.

Zur Auswertung wurden zwei Kennwerte der absoluten Genauigkeit ausgewählt, der Absolute Accuracy Index und der Bias. Der Bias gibt die Richtung der Fehler bei der Einschätzung an, d. h. es liegt eine Tendenz zur Unter- oder Überschätzung vor (Schraw 2009). Dabei werden die Differenzen zwischen den Urteilen und den Leistungen aufsummiert und durch die Anzahl der Basiskompetenzen geteilt. Liegt die Einschätzung über der Leistung, so gibt der Bias einen Wert größer Null wieder, d. h. der Lernen-

de hat eine Tendenz seine Leistung zu überschätzen. Liegt die Einschätzung unter der Leistung, so gibt der Bias einen Wert kleiner Null wieder, d. h. der Lernende hat eine Tendenz seine Leistung zu unterschätzen. Gibt es keine Differenzen, so wird der Wert Null wieder gegeben. Der Absolute Accuracy Index gibt an, inwieweit die Einschätzung von der Leistung abweicht (Schraw 2009). Hierzu werden die quadrierten Differenzen zwischen Urteil und Leistung zu jeder Basiskompetenz aufsummiert und durch die Anzahl der Basiskompetenzen geteilt. Gibt es keine Differenzen zwischen den abgegebenen Urteilen und der erhobenen Leistung, so hat der AAI den Wert 0. Liegen die Urteile und die Leistung eine Stufe auseinander im Mittel über alle Basiskompetenzen, so hat der AAI den Wert 0,11. Liegen Urteile und Leistung 2 Kategorien auseinander, so hat der AAI den Wert 0,44. Dies sind lediglich Richtwerte, da die Differenzen sich bei den Basiskompetenzen unterscheiden können.

Ergebnisse

Bei einer Klassifizierung der Stichprobe nach dem Geschlecht ließen sich nur geringe Unterschiede zwischen den zwei Gruppen erkennen. Unterschiede in Bezug auf eine Unterschätzung bei Mädchen konnten nicht beobachtet werden. In der vorliegenden Stichprobe überschätzten sich beide Gruppen.

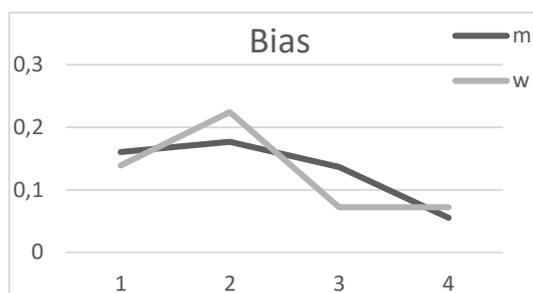


Abbildung 1: Bias nach Geschlechtern

In der Forschung zu metakognitiven Urteilen konnte in einigen Studien ein Zusammenhang zwischen dem Level des eigenen Könnens und der Fähigkeit, sich selbst adäquat einschätzen zu können, nachgewiesen werden (Bol, Hacker 2012). Daher wurde eine Klassifizierung nach der Entwicklung der Testleistung vorgenommen. Das erste Cluster beinhaltet ungefähr die Hälfte der Stichprobe und war bezüglich der Testleistung weitestgehend stabil im unteren Bereich. Das zweite Cluster beinhaltet ungefähr ein Viertel der Lernenden und zeichnet sich durch eine positive Entwicklung der Testleistung aus. Das dritte Cluster beinhaltet ebenfalls ungefähr ein Viertel der Lernenden, welche zu allen Messzeitpunkten eine gute Testleistung zeigten.

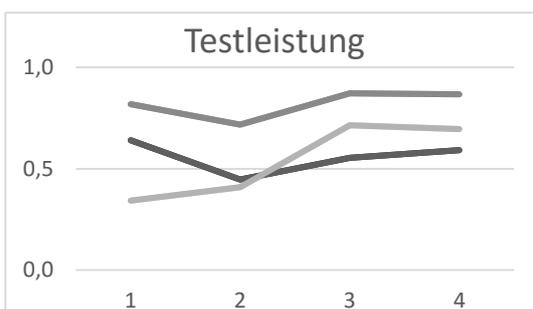


Abbildung 2: Entwicklung der Testleistung in den Clustern

Diese drei Cluster zeigen auch bei der Entwicklung des Bias und des AAI unterschiedliche Entwicklungen. Das dritte Cluster mit der guten Testleistung zeigt nur eine geringe Tendenz zur Überschätzung und im Vergleich zu den anderen Clustern eine bessere Genauigkeit bei der Einschätzung. Das zweite Cluster kann neben der Steigerung der Testleistung auch eine Steigerung der Genauigkeit zeigen, aber vor allem eine Abnahme der Überschätzung. Das erste Cluster zeigt ähnlich wie bei der Testleistung auch beim Bias und AAI nur geringe Veränderungen und keine erkennbaren Veränderungen.

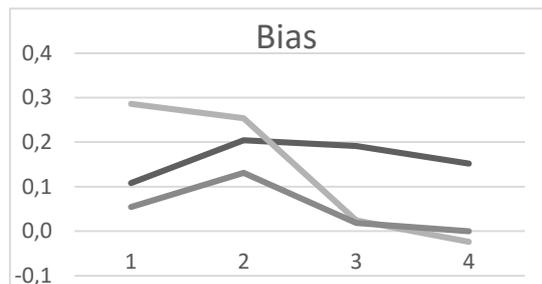


Abbildung 3: Bias nach Clustern

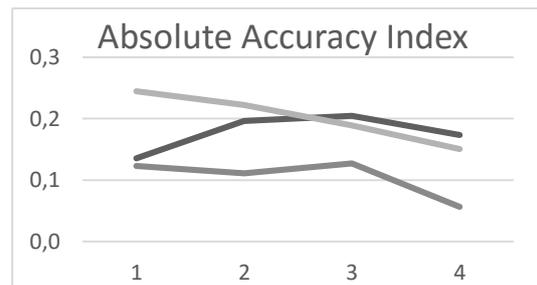


Abbildung 4: AAI nach Clustern

Diskussion und Ausblick

Aus den erhobenen Daten ist zu erkennen, dass ein Teil der Lernenden ohne zusätzliche Hilfestellungen in der Lage ist, den Übungsbedarf zu bestimmen oder zumindest durch Übung sich darin zu verbessern. Es kann daher davon ausgegangen werden, dass ein Teil der Lernenden produktiv mit den Selbstdiagnosebögen arbeiten kann.

Es ist aber auch erkennbar, dass ein Teil der Lernenden neben Selbstdiagnosebögen weitere Hilfestellungen zum Verbessern der Genauigkeit benötigt. Eventuell kann diese Hilfestellung durch den sozialen Vergleich, der beispielsweise durch Partnerdiagnose (Reiff 2006) erfolgen kann, gegeben werden.

Literatur

- Achilles, H. (2011). Selbst organisierte Prüfungsvorbereitung mithilfe von Selbsteinschätzungsbogen unterstützen. *PM. Praxis der Mathematik in der Schule*, 41, 17–22.
- Bol, Linda; Hacker, Douglas J. (2012): Calibration research: where do we go from here? In: *Frontiers in psychology*, 3, 229
- Reiff, R. (2006). Selbst- und Partnerdiagnose im Mathematikunterricht. Gezielte Förderung mit Diagnosebögen. *Friedrich Jahresheft*, 24, 68–72.
- Schraw, Gregory (2009): Measuring Metacognitive Judgements. In: Douglas J. Hacker, John Dunlosky und Arthur C. Graesser (Hg.): *Handbook of metacognition in education*. New York: Routledge, 415–429.