Marcel SCHAUB, Darmstadt

Einsatz des Elementarisierenden Testens im Ein- und Ausgangstest des online-Vorkurses VEMINT

Aufgrund hoher Abbruchquoten in mathematikaffinen Studiengängen und starker Heterogenität der Studienanfänger_innen werden an fast jeder Hochschule Vor- bzw. Brückenkurse angeboten. Um insbesondere in online-Vorkursen mit großer inhaltlicher Breite individuelle Lernpfade und Lernzielbildungen zu unterstützen, ist eine (digitale) Diagnose des aktuellen Lernstands notwendig, mit der Defizite lokalisiert werden können.

Elementarisierendes Testen

Für das Prüfen von Mathematischem Grundwissen und Grundkönnen (Bruder et al., 2015) stellt Feldt-Caesar (2017) folgende Dialektik fest: Die zu prüfenden Inhalte sollten möglichst verknüpft geprüft werden, aber auch eine präzise Diagnostik ermöglichen. Mit dem elementarisierenden Testen (ebd.) können beide Ziele zeitökonomisch erreicht werden. In diesem adaptiven Verfahren (ebd.) werden alle zu prüfenden Inhalte des Tests in (teilweise) mehrschrittigen Aufgaben innerhalb der Hauptlinie gestellt. Beantwortet man eine solche Aufgabe falsch, wird man in eine elementarisierende Schleife gelenkt, die Teilanforderungen dieser Hauptlinienaufgabe isoliert und im Sinne des Elementarisierten Testens abbildet und in der Hauptlinie wieder mündet. Werden alle Aufgaben einer Schleife korrekt beantwortet und somit kein Defizit lokalisiert, wird am Ende des Tests eine Parallelaufgabe zur entsprechenden Hauptlinienaufgabe gestellt.

Die Grenzen des Verfahrens lassen sich neben den allgemeinen Grenzen digitalen Testens durch zwei Aspekte beschreiben: Zum einen prüfen die Schleifenaufgaben nur notwendige Bedingungen, aber keinesfalls hinreichende Bedingungen zum Lösen der Hauptlinienaufgabe (ebd.). Zum anderen kann die Isolierung der Stoffelemente aus der komplexen Aufgabe zu einer punktuellen Erhöhung der Anforderung führen (ebd.). Ziel der Pilotstudie über den Ein- und Ausgangstest des online-Vorkurs VEMINT in Darmstadt im WS 17/18 war es die eingesetzten Schleifen zu evaluieren und typische Fehlerphänomene zu identifizieren.

Um die Qualität des Diagnoseverfahrens messen und unterschiedliche Varianten hinsichtlich ihres diagnostischen Potentials untersuchen zu können, hat Feldt-Caesar (2017) unterschiedliche Quoten definiert, die hier modifiziert dargestellt werden:

Unter der *Fehleraufklärungsquote* wird das Verhältnis aus dem Anteil der Lernenden, deren Fehler zumindest teilweise aufgeklärt werden konnte, zum

Anteil der Lernenden, die einen Fehler in der Aufgabe produziert haben, verstanden.

Als *Kritische Gruppe* wird das Verhältnis aus dem Anteil der Lernenden, die alle Schleifenaufgaben korrekt, aber die Parallelaufgabe falsch gelöst haben, und dem Anteil der Lernenden, die einen Fehler in der Hauptlinienaufgabe produziert haben, bezeichnet.

Das diagnostische Potential einer elementarisierenden Schleife ist umso größer, je höher die Fehleraufklärungsquote und je geringer der Anteil der Kritischen Gruppe ist (ebd.). Mit Hilfe dieser Kennzahlen kann außerdem die Wirkung von optimierenden Maßnahmen eingeschätzt werden.

Empirische Ergebnisse aus der Pilotstudie im WS 17/18

Im Eingangstest des Vorkurses wurden vier elementarisierende Schleifen zu den Themen Termumformungen sowie zur Differential- und Integralrechnung eingesetzt und für N=272 untersucht.

Der Graph der Funktion f mit $f(x) = 3x^2 - 6x - 9$ schließt mit der x-Achse eine Fläche vollständig ein. Bestimmen Sie den Flächeninhalt dieser Fläche.

Abb. 1: Aufgabe "Flächeninhalt"

Feldt-Caesar (2017) zeigte bereits positive Effekte auf die Fehleraufklärungsquote durch den Einsatz von Distraktoren mit diagnostischem Potential (Winter, 2011). In dieser Studie wurden weitere typische Fehlerphänomene identifiziert und deren möglichen Ursachen teilweise durch diagnostische Interviews mit Erstsemesterstudierenden aufgeklärt. Zu der Aufgabe Flächeninhalt (bei $N = 272$ Eingaben gab es $N_F = 139$ Falschantworten, siehe Abbildung 1) konnten drei typische Fehlerphänomene identifiziert werden (siehe Abbildung 2). Bei der Aufgabe zur Tangentenbestimmung (siehe Abbildung 3) konnte ein typisches Fehlerphänomen mit $t(x) = -8x + 17$, das auf die fehlerhafte Multiplikation

Typisches Fehlerphänomen	**Anteil an allen Fehlern**	**Mögliche Ursache**
-32	12,2%	Das Integral wurde korrekt bestimmt und berechnet, jedoch das Ergebnis nicht im Betrag gesetzt (bereits bei ebd.)
22	5,8%	Bei der Berechnung des bestimmten Integrals wurden die Funktionswerte der Stammfunktion addiert.
$x^3 - 3x^2 - 9x$	5,0%	Es wurde die Stammfunktion gebildet.

Abb. 2: Typische Fehlerphänomene zur Aufgabe "Flächeninhalt"

Gegeben ist die Funktionsgleichung einer Funktion f durch: $f(x) = 2x^2 - 7$

Bestimmen Sie die Funktionsgleichung der Tangente t an den Graphen der Funktion f im Punkt $P(-2|1)$.

Abb. 3: Aufgabe "Tangentenbestimmung"

zweier negativer Zahlen bei der Bestimmung des y-Achsenabschnitts zurückzuführen ist (7,3% aller gemachter Fehler bei $N_F = 109$ Falschantworten und $N = 254$ Eingaben). Es konnte sogar eine ganze Antwortklasse gefunden werden: Alle Lösungen mit korrekt bestimmter Steigung und falsch bestimmtem y-Achsenabschnitt (21,2% aller Fehler). Bei allen Eingaben kann das Aufstellen einer linearen Funktion als Defizit zurückgemeldet werden. Mit Hilfe von STACK, das mit einem CAS arbeitet, sind nicht nur Aufgaben im offenen Format und die Identifikation von typischen Fehlern möglich (Kallweit et al., 2017), sondern es können ganze Antwortklassen geprüft werden. Betrachtet man das fiktive Szenario retrospektiv, dass diese Antwortklasse bereits als aufgeklärt identifiziert geworden wären, würde die Fehleraufklärungsquote für diese Aufgabe von 62,2% auf 77,1% steigen sowie die Kritische Gruppe von 14,7% auf 9,2% sinken. Somit kann in zukünftigen Durchläufen das diagnostische Potential dieser Schleife deutlich erhöht werden und die Verwendung von STACK mit typischen Fehlerphänomenen und Antwortklassen dient als ergänzendes Mittel.

Ein weiterer Vorteil ist, dass Lernende, die einen typischen Fehler oder eine Antwort, die sich einer Antwortklasse zuordnen lässt, produziert haben, nicht in die Schleife geleitet werden müssen, da der Fehler bereits aufgeklärt ist. Somit wird das Verfahren zeitökonomischer. Eine weitere Möglichkeit, die beispielsweise bei Termumformungen relevant wird, ist, dass Lernende, die ein Zwischenergebnis eingeben, das zwar korrekt ist, aber nicht zu Ende geführt wurde, nicht die ganze Schleife durchlaufen müssen, sondern nur noch diejenigen, über deren Stoffelemente noch keine Rückschlüsse gezogen werden können.

Neben der Zielstellung die Fehlerursachen zu den typischen Fehlerphänomenen zu identifizieren wurden diagnostische Interviews zur Kritischen Gruppe in der Mathematiklehrveranstaltung für Bauingenieure im ersten Semester durchgeführt. Analysen der Interviews deuten an, dass Lernende, die alle Schleifenaufgaben korrekt lösen, oft keinen Handlungsplan für die komplexe Aufgabe aufstellen können. Gezeigt hat sich insbesondere, dass das isolierte Prüfen von Zusammenhängen im MC-Format die Anforderung stark heruntersetzt, da der Zusammenhang nicht mehr selbstständig realisiert werden muss, sondern (ggf. durch Ausschlussverfahren) identifiziert werden kann. Bei Aufgaben zum Vereinfachen von Termen hat eine Studentin die Kenntnisse nicht auf die komplexe Aufgabe anwenden können, da die Aufgabe „zu komplex“ sei, auch wenn der Ausdruck nach vorigen Umformungen fast identisch komplex zum Ausdruck in der Schleife ist. Hier lässt sich die Vermutung aufstellen, dass die regelirrelevanten Merkmale der komplexen Ausdrücke das Erkennen der Notwendigkeit einer Regel verhindern. Eine weitere Aussage deutet außerdem daraufhin, dass die Erwartungshaltung zur Gestaltung an einen solchen Test die Motivation und Konzentration beeinflusst: So erwartete die betroffene Person, dass zu jedem Aufgabentyp (bezogen auf Inhalte) nur eine in einem solchen Test gestellt wird und war deshalb demotiviert spätere Aufgaben der Schleifen konzentriert zu bearbeiten. Somit gilt es in Zukunft auch die Akzeptanz des verwendeten Verfahrens zu prüfen.

Ausblick

Im kommenden Semester werden die Optimierungen implementiert und mit einem neu entwickelten Plugin (https://github.com/LG3696/ddtaquiz) erneut in moodle umgesetzt. Mit Hilfe des Plugins ist eine verbesserte Adaption des Elementarisierenden Testens möglich sowie der Einsatz von adaptiven Feedbackeinheiten. Ziel für das WS 18/19 ist die Validierung der typischen Fehler, die Untersuchung der Akzeptanz des Diagnoseverfahrens und weiterführende Untersuchungen zur Kritischen Gruppe.

Literatur

Bruder, R.; Feldt-Caesar, N.; Pallack, A.; Pinkernell, G.; Wynands, A. (2015): Mathematisches Grundwissen und Grundkönnen in der Sekundarstufe II. In: W. Blum, S. Vogel, C. Drüke-Noe und A. Roppelt (Hg.): Bildungsstandards aktuell. Mathematik in der Sekundarstufe II. Braunschweig: Schroedel, S. 108–124.

Feldt-Caesar, N. (2017): Konzeptualisierung und Diagnose von mathematischem Grundwissen und Grundkönnen. Eine theoretische Betrachtung und exemplarische Konkretisierung am Ende der Sekundarstufe II. Wiesbaden: Springer Fachmedien.

Kallweit, M.; Schaub, M.; Feldt-Caesar, N.; Bruder, R.; Krusekamp, S.; Neugebauer, C.; Winter, K. (2017): Digitale Diagnostische Testaufgaben – Theoretisches Design und interaktives Beispiel. In: Beiträge zum Mathematikunterricht 2017, S. 1389–1392. Münster: WTM.

Winter, Kathrin (2011): Entwicklung von Item-Distraktoren mit diagnostischem Potential zur individuellen Defizit- und Fehleranalyse. Didaktische Überlegungen, empirische Untersuchungen und konzeptionelle Entwicklung für ein internetbasiertes Mathematik-Self-Assessment. Münster: WTM.