

Diagnosequalität von Lehramtsstudierenden – Können sie Probleme von Schüler:innen beim Debugging erfassen?

Heike Wachter ¹ und Tilman Michaeli ¹

Abstract: Lehrkräfte müssen Schüler:innen beim Debugging unterstützen können. Angehende Lehrkräfte haben jedoch oft Schwierigkeiten mit der Diagnose von Schülerproblemen. Entsprechende Diagnosefähigkeiten sollten daher bereits in der Lehramtsausbildung gefördert werden. Jedoch fehlt es bisher an Erkenntnissen, welche Diagnosefähigkeiten Lehramtsstudierende für diese Unterrichtssituationen besitzen und wie diese in der Lehrkräfteausbildung gefördert werden können. In diesem Beitrag untersuchen wir daher anhand von geskripteten Videovignetten, die Diagnosefähigkeiten von Lehramtsstudierenden. Die Ergebnisse zeigen eine geringe Genauigkeit bei der Diagnosequalität, was Hinweise auf Förderbedarf in der Lehramtsausbildung gibt.

Keywords: Debugging, Diagnosefähigkeit, Videovignetten, Lehramtsstudierende, Diagnosequalität

1 Einleitung

Debugging im Informatikunterricht ist für Schüler:innen und Lehrkräfte eine Herausforderung. Insbesondere Programmieranfänger:innen haben oft große Schwierigkeiten Fehler im Programmcode zu finden und zu beheben. Viele Schüler:innen sind daher auf die Unterstützung der Lehrkraft angewiesen. Häufig führt das dazu, dass die Lehrkraft von Lernenden zu Lernenden eilt, um Hilfestellung zu geben [MR19]. Während der Unterstützung der Schüler:innen durchläuft die Lehrkraft einen mehrstufigen Diagnose- und Interventionsprozess [HM23]. Zunächst muss die Lehrkraft diagnostizieren, weshalb der oder die Schüler:in nicht in der Lage ist, das Problem selbstständig zu lösen. Anschließend muss sie mit einer geeigneten Intervention reagieren. Allerdings haben angehende Lehrkräfte zu Beginn ihrer Lehrtätigkeit oftmals Schwierigkeiten mit dem Diagnostizieren von Unterrichtssituationen [CSea20]. Deshalb sollten Diagnosefähigkeiten bereits in der Lehrkräfteausbildung gefördert werden. Bisher sind Diagnosefähigkeiten beim Debugging jedoch weder üblicher Inhalt der informatikdidaktischen Lehramtsausbildung noch wissenschaftlich untersucht. Um Diagnosefähigkeiten gezielt fördern zu können, wird daher in diesem Beitrag die Qualität der Diagnose von Lehramtsstudierenden für Debuggingsituationen untersucht.

2 Hintergrund und Forschungsstand

Im Unterrichtsalltag werden Lehrkräfte häufig mit Situationen konfrontiert, in denen sie die unterschiedlichen Lernbedürfnisse, Fähigkeiten, Interessen und Motivationen der

¹ TU München, School of Social Sciences and Technology, Professur für Didaktik der Informatik, Arcisstraße 21, 80333 München, Heike.Wachter@tum.de,  <https://orcid.org/0000-0002-5090-7183>;
Tilman.Michaeli@tum.de,  <https://orcid.org/0000-0002-5453-8581>

Schüler:innen diagnostizieren müssen [VR09]. Diagnosefähigkeiten, also das Beurteilen relevanter Merkmale für das Lernen von Schüler:innen [BKea18], sind somit eine zentrale Kompetenz von Lehrkräften. In der Informatikdidaktik wurden Diagnosefähigkeiten von Lehrkräften bisher kaum untersucht. Für debuggingspezifische Unterrichtssituationen wurde ein Prozessmodell für Diagnose und Intervention beim Debugging [HM23] entwickelt, welches die Schritte des Diagnoseprozesses abbildet und aufzeigt, welche Kenntnisse jeweils benötigt werden.

Zur Untersuchung von Diagnosefähigkeiten und für die Förderung in der Lehrkräfteausbildung werden häufig Videovignetten verwendet, beispielsweise in der Mathe- [SCea23] oder Physikdidaktik [KF25]. Videovignetten haben den Vorteil, dass sie die Komplexität der Realität im Sinne repräsentationalen Scaffoldings reduzieren [Fea22] und dadurch die Beobachtung des spezifischen Lehr-Lern-Verhaltens erleichtern. Die Qualität einer Diagnose lässt sich mit Hilfe der diagnostischen Genauigkeit (Accuracy) operationalisieren [CHea22]. Die Genauigkeit gibt an, wie exakt die Beurteilung des Wissens und der Kompetenz eines oder einer Schüler:in durch die Lehrkraft ist. Codreanu et al. [CSea20] erhoben zur Messung der Genauigkeit verschiedene Items zu mathematischen Inhaltsdimensionen wie mathematisches Wissen, Methodenwissen und Problemlösungsstrategien mit einer vierstufigen Likert-Skala. Schons et al. [SOea24] verwendeten ebenfalls ein geschlossenes Antwortformat, bei dem die korrekte Diagnose aus einer Liste von acht Aussagen ausgewählt werden musste. In beiden Studien wurde die Qualität der Diagnose mit einem Punktesystem bewertet. Kramer et al. [KFea21] nutzten dagegen Aufgabenstellungen mit offenem Antwortformat. Für die Genauigkeit der Diagnose wurde ebenfalls ein Punktesystem angewandt, bei dem 0 Punkte bei sehr ungenauen Antworten und 1-3 Punkte für Antworten mit besserer Qualität vergeben wurden. Zusätzlich bewerteten sie die Genauigkeit durch den Vergleich der individuellen Beobachtung der Lehrkräfte mit den in die Videos eingebetteten, geskripteten Herausforderungen. In der Informatikdidaktik wurde die Qualität von Diagnosen jedoch bisher nicht untersucht. Lediglich Aspekte, auf die erfahrene Lehrkräfte bei der Diagnose von Debuggingsituationen achten, wurden identifiziert [WM24]. Sie fanden verschiedene situationspezifische Aspekte, die sich in die Ebenen Konzeptwissen, Problemlösestrategien und emotional-motivationale Aspekte einordnen ließen.

3 Vorgehen

Diese Arbeit untersucht die Qualität der Diagnose von Lehramtsstudierenden bei der Diagnose von Schüler:innen beim Debugging. Folgende Forschungsfrage wird dabei adressiert: *Wie ist die Qualität der Diagnose von Lehramtsstudierenden beim Diagnostizieren von Schüler:innen beim Debugging?* Zur Beantwortung der Forschungsfrage verwendeten wir vier geskriptete Videovignetten, die typische Debuggingsituationen für zwei Schülerinnen im Klassenzimmer zeigen. Dies ermöglicht uns die Diagnosefähigkeit der Lehramtsstudierenden in einer bestimmten Situation zu untersuchen. Die einzelnen Situationen basieren auf der Literatur und Unterrichtspraxis [ZMR23]. Die Videovignetten zeigen jeweils das

Verhalten, die Kommunikation und den Inhalt des Bildschirms von zwei Schüler:innen, während sie mit einem bestimmten Fehler konfrontiert sind (vgl. Abb. 1). Bei der Bearbeitung der Videovignetten müssen die Studierenden verschiedene situationsspezifische Aspekte im Verhalten der Schüler:innen wahrnehmen und bewerten, um zu einer korrekten Diagnose zu gelangen. Die zu diagnostizierenden Aspekte können verschiedenen Ebenen zugeordnet werden (Konzeptwissen, Problemlösestrategien und emotional-motivationale Aspekte). In der ersten Vignette fehlt den Schüler:innen beispielsweise Wissen bzgl. der Syntax (Konzeptwissen). Außerdem zeigen sie ein Trial-and-Error-Verhalten (Problemlösestrategien) und sind schnell frustriert (emotional-motivational). Um diese Aspekte diagnostizieren zu können, enthält jede Videovignette entsprechende *Cues* (Hinweise). So setzen die Schüler:innen zunächst eine fehlende Klammer an die falsche Stelle, was entsprechend auf fehlendes syntaktisches Wissen hinweist. Außerdem äußern sie explizit ihre Frustration. Cues für das Trial-and-Error-Verhalten finden sich sowohl in der unsystematischen Navigation in der Programmierumgebung, als auch in der Kommunikation der Schüler:innen.

Datenerhebung: Die Datenerhebung erfolgte in einer Einführungsvorlesung in die Informatikdidaktik an einer deutschen Universität. Die Teilnehmenden waren Studierende am Ende des Bachelor-Studiums bzw. am Anfang des Master-Studiums Lehramt Informatik. Innerhalb der Veranstaltung, bearbeiteten sie selbstständig Diagnoseaufgaben mit unterschiedlichem Schwierigkeitsgrad zu den Videovignetten in einer Lernumgebung. Zwei der Diagnoseaufgaben waren unstrukturiert, d.h. es gab lediglich ein Textfeld in das die Diagnose eingetragen werden sollte. Für die beiden anderen Vignetten wurde das Antwortformat stärker strukturiert, d.h. die Lehramtsstudierenden wurden durch Fragestellungen, wie etwa: „Beschreibe, welche Aspekte des Verhaltens der Schülerinnen du in der gezeigten Unterrichtssequenz für diagnostisch relevant hältst. Diese Aspekte können sich sowohl auf das Konzeptwissen der Schülerinnen, ihren Problemlösungsprozess als auch auf affektive Aspekte wie Motivation und Emotion beziehen.“ schrittweise zu einer Diagnose geleitet. Genauso wurde explizit nach dem zugrunde liegenden fachlichen Fehler gefragt. Alle Aufgabenstellungen hatten ein offenes Antwortformat. Insgesamt nahmen neun Lehramtsstudierende (4 m, 5 w) an der Studie teil. Die Studierenden in unserer Stichprobe hatten eher wenig Unterrichtserfahrung, die kaum über die Pflichtpraktika hinausging. Da nicht alle Teilnehmer:innen aufgrund der zur Verfügung stehenden Zeit alle Aufgaben bearbeitet haben, erhielten wir insgesamt 6 vollständige Datensätze und 3 Datensätze bei denen die letzte Aufgabe fehlte.

Auswertung: Die Analyse der Diagnosequalität erfolgte über eine deduktive qualitative Inhaltsanalyse nach Mayring. Angelehnt an Kramer et al. [KFea21] haben wir die Genauigkeit durch den Vergleich der individuellen Beobachtung der Studierenden mit den in die Videos eingebetteten Aspekten verglichen. Dafür wurde für jede Diagnose codiert, ob die Aspekte für Konzeptwissen, Problemlösestrategie und Motivation/Emotion sowie der fachliche Fehler vollständig, teilweise oder gar nicht erfasst wurden. Wir bewerteten die Antworten in jeder Kategorie mit Punkten: vollständig (2 Punkte), teilweise (1 Punkt) oder gar nicht erfasst (0 Punkte). Die maximale Punktzahl, die für eine Diagnose erreicht werden konnte, war somit 8 Punkte.

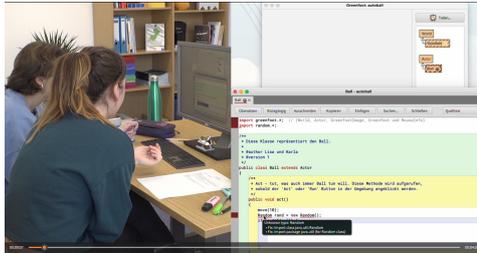


Abb. 1: Screenshot einer Vignette

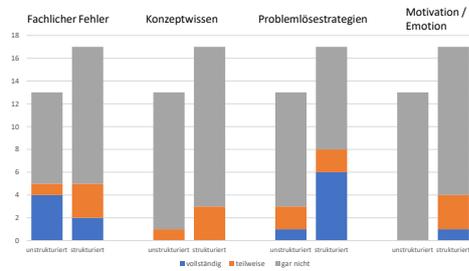


Abb. 2: Ergebnisse der Analyse

4 Ergebnisse

Die Auswertung der Diagnosequalität zeigt, dass die Gesamtpunktzahl der Studierenden und somit die Qualität der Diagnose eher gering war. Durchschnittlich erreichten die Studierenden höchstens 6 (strukturiert) bzw. 3 (unstrukturiert) Punkte. Es gab jedoch auch finale Bewertungen von 0 Punkten. In Abb. 2 werden die Ergebnisse der Datenanalyse dargestellt. Der Unterschied zwischen unstrukturiert und strukturiert (Schnitt 1,6 vs. 3,2) ist deutlich zu erkennen. Die zu diagnostizierende Aspekte wurden bei der strukturierten Aufgabenstellung für alle Bereiche besser erfasst als bei der Unstrukturierten. Allerdings zeigt sich auch bei den strukturierten Aufgabenstellungen eine eher geringe Diagnosequalität, da der Anteil der vollständig und teilweise erfassten Aspekte an der Gesamtheit eher gering ist.

Der *fachliche Fehler* wurde von vielen Studierenden in ihren Diagnosen nicht genannt. Die Ursache hierfür ist unklar, entweder hielten sie ihn nicht für relevant oder konnten ihn nicht identifizieren. Vergleicht man die Ergebnisse von unstrukturierter und strukturierter Aufgabenstellung fällt auf, dass der fachliche Fehler die einzige Kategorie ist, bei der sich keine Steigerung durch die strukturierte Aufgabenstellung gezeigt hat – sondern im Gegenteil der fachliche Fehler bei der unstrukturierten Aufgabenstellung häufiger vollständig erfasst wurde. Obwohl auch bei der strukturierten Aufgabenstellung explizit nach dem fachlichen Fehler gefragt wurde, scheint sich hier die Aufmerksamkeit auf andere Bereiche verschoben zu haben. Auch das *Konzeptwissen* wurde selten in den Diagnosen adressiert – und wenn dann oftmals nur sehr ungenau. Beispielsweise wurde in einer Diagnose „fehlendes Fachwissen“ geantwortet, jedoch nicht benannt welches Konzeptwissen den Schülerinnen fehlt. In vielen Fällen wurden aber auch andere Aspekte der Vignette genannt wie z.B., dass den Schülerinnen Englischkenntnisse fehlen. Diese Aspekte sind zwar in der Vignette dargestellt und damit nicht falsch, jedoch sind sie für eine korrekte Diagnose nicht zielführend. Das fehlende Konzeptwissen wurde zwar bei der strukturierten Aufgabenstellung etwas häufiger erfasst. Jedoch wurde es in beiden Fällen nie vollständig erfasst. In vielen Diagnosen der Studierenden wurden auch keine Aspekte zur Kategorie *Problemlösestrategien* genannt. Wurden sie jedoch erfasst, waren sie häufig korrekt und vollständig dargestellt. Im Vergleich zu Aspekten der Kategorien Konzeptwissen und

Motivation/Emotion wurden die Aspekte zur Problemlösestrategien am häufigsten erfasst. In dieser Kategorie zeigen sich auch deutlich bessere Ergebnisse bei den strukturierten Aufgabenstellungen gegenüber den unstrukturierten. Auf *motivationale und emotionale Aspekte* wurde dagegen selten eingegangen. Bei der Kategorie Motivation/Emotion fällt auf, dass sie sogar nur bei der strukturierten Aufgabenstellung überhaupt erfasst wurde, vermutlich aufgrund der expliziten Frage nach diesen Aspekten in der Aufgabenstellung. Abschließend fällt auf, dass in vielen Diagnosen jeweils nur eine der vier Kategorien adressiert wurde.

Insgesamt wurde bei den Diagnosen häufig auf Aspekte geachtet, die wir als wenig relevant erachten. Zudem waren die Diagnosen oft ungenau. Ein Beispiel dafür ist die allgemeine Diagnose „SuS verstehen die Fehler nicht + fehlendes Vorwissen“, die keinen Bezug zur konkreten Situation hatte. In einigen Fällen wurde ein Cue zwar wahrgenommen, aber falsch interpretiert, wie bei der Aussage „Ihr Code hat eine Exception ausgelöst, sie können die Fehlermeldung nicht wegeklicken“. Eine korrekte Interpretation wäre gewesen, dass die Schülerinnen nicht motiviert sind, die Fehlermeldung zu bearbeiten oder das Lesen der Fehlermeldung nicht als Schritt zur Lösung auffassen – und deshalb wiederholt versuchen die Null-Pointer-Exception zu schließen.

5 Diskussion und Fazit

Die Ergebnisse der Untersuchung zeigen insgesamt eine geringe Genauigkeit der Diagnose. Häufig wurden dabei Aspekte genannt, die für die Diagnose nicht relevant waren. Diagnosefähigkeiten für das Diagnostizieren von Schüler:innen beim Debugging sind bisher kein typischer Inhalt der informatikdidaktischen Ausbildung. Insofern ist das Abschneiden der Lehramtsstudierenden nur bedingt überraschend, zeigt aber einen deutlichen Bedarf auf, Diagnosefähigkeiten bereits im Studium gezielt zu fördern.

Die Auswertung der Diagnoseaufgaben deutet auf Unterschiede in der Qualität der Diagnosen bei strukturierten Aufgaben im Vergleich zu unstrukturierten Aufgaben hin. Es scheint den angehenden Lehrkräften leichter zu fallen, eine Diagnose zu erstellen, wenn sie durch Hilfestellungen, hier gezielte Fragen, durch den Diagnoseprozess geführt werden. Dies zeigt einen ersten Ansatz zur Förderung der Diagnosefähigkeiten von Lehramtsstudierenden auf.

Eine Limitation dieser Arbeit ist die kleine Stichprobe, die die Aussagekraft der Ergebnisse in Bezug auf ihre Verallgemeinerbarkeit einschränkt. Jedoch geben die Ergebnisse erste Hinweise auf Förderpotenziale bei angehenden Lehrkräften und dienen als Grundlage für weitere Untersuchungen.

Zusammenfassend tragen unsere Ergebnisse zu einem besseren Verständnis von Diagnosefähigkeiten von Lehramtsstudierenden beim Diagnostizieren von Schüler:innen beim Debugging bei, die bisher innerhalb der Informatikdidaktik nicht untersucht wurden. Neben der Methodik zur Erfassung der Diagnosequalität zeigen unsere Ergebnisse nicht nur

einen dringenden Förderbedarf, sondern geben auch erste Hinweise auf Probleme die Lehramtsstudierende bei der Diagnose in Unterrichtssituationen haben.

Literaturverzeichnis

- [BKea18] Binder, Karin; Krauss, Stefan; et al.: Diagnostic skills of mathematics teachers in the COACTIV study. Diagnostic competence of mathematics teachers: Unpacking a complex construct in teacher education and teacher practice, 2018.
- [CHea22] Chernikova, Olga; Heitzmann, Nicole; et al.: A theoretical framework for fostering diagnostic competences with simulations in higher education. Learning to Diagnose with Simulations, 2022.
- [CSea20] Codreanu, Elias; Sommerhoff, Daniel; et al.: Between authenticity and cognitive demand: Finding a balance in designing a video-based simulation in the context of mathematics teacher education. Teaching and Teacher Education, 2020.
- [Fea22] Fischer, Frank; et al.: Representational scaffolding in digital simulations—learning professional practices in higher education. Information and Learning Sciences, 2022.
- [HM23] Hennig, Heike; Michaeli, Tilman: Entwicklung eines Prozessmodells für Diagnose- und Intervention von Lehrkräften beim Debugging. In: INFOS 2023. Gesellschaft für Informatik eV, 2023.
- [KF25] Krumphals, Ingrid; Feser, Markus Sebastian: Fostering pre-service physics teachers' diagnostic skills and readiness through video vignettes and micro-teaching sessions: An exploratory single-case study. In: Journal of Physics: Conference Series. 2025.
- [KFea21] Kramer, Maria; Förtsch, Christian; et al.: Investigating pre-service biology teachers' diagnostic competences: Relationships between professional knowledge, diagnostic activities, and diagnostic accuracy. Education Sciences, 2021.
- [MR19] Michaeli, Tilman; Romeike, Ralf: Current status and perspectives of debugging in the k12 classroom: A qualitative study. In: 2019 IEEE EDUCON. IEEE, 2019.
- [SCea23] Sommerhoff, Daniel; Codreanu, Elias; et al.: Pre-service teachers' learning of diagnostic skills in a video-based simulation: Effects of conceptual vs. interconnecting prompts on judgment accuracy and the diagnostic process. Learning and Instruction, 2023.
- [SOea24] Schons, Christian; Obersteiner, Andreas; et al.: Toward adaptive support of pre-service teachers' assessment competencies: Log data in a digital simulation reveal engagement modes. Learning and Instruction, 2024.
- [VR09] Vogt, Franziska; Rogalla, Marion: Developing adaptive teaching competency through coaching. Teaching and teacher education, 2009.
- [WM24] Wachter, Heike; Michaeli, Tilman: Analyzing teachers' diagnostic and intervention processes in debugging using video vignettes. In: International Conference on Informatics in Schools: Situation, Evolution, and Perspectives. Springer, 2024.
- [ZMR23] Zoppke, Till; Michaeli, Tilman; Romeike, Ralf: Individuelle Unterstützung beim Debuggen—Video-Vignetten für die Lehrkräftebildung. In: INFOS 2023. Gesellschaft für Informatik eV, 2023.