

Verena Schürmann¹, Theresa Gruyters², Songül Gözümlü³, Daniel Bodemer⁴ & Nicki Marquardt⁵

Verhaltensverankerte Einstufungsskalen (BARS) zur Erfassung studentischer Kollaboration

Zusammenfassung

Hochschulen stehen in der Verantwortung, Studierende effektiv darauf vorzubereiten, komplexe Probleme und Herausforderungen in transdisziplinären Teams zu lösen. Der vorliegende Beitrag diskutiert das Potenzial von verhaltensverankerten Einstufungsskalen (BARS) für die Messung und Förderung von Kollaboration im Hochschulkontext. In einem iterativen Verfahren wurden zehn verhaltensverankerte Skalen erstellt. Diese wurden unter Einbeziehung verschiedener Expertengruppen (Forschende, Lehrende, Studierende) evaluiert. Die Ergebnisse legen nahe, dass die Verhaltensanker insgesamt repräsentativ und verständlich sind. Ferner werden Anwendungsmöglichkeiten der BARS im Hochschulkontext aufgezeigt.

Schlüsselwörter

Kollaboration, Hochschule, Verhaltensverankerte Einstufungsskalen (BARS), Instrument

-
- 1 Corresponding Author; Hochschule Rhein-Waal, Universität Duisburg-Essen; verena.schuermann@hsrw.eu; ORCID 0000-0001-8008-1215
 - 2 Hochschule Rhein-Waal; theresa.gruyters@hsrw.org
 - 3 Hochschule Rhein-Waal; songuel.goezuem@hsrw.org
 - 4 Universität Duisburg-Essen; bodemer@uni-due.de; ORCID 0000-0003-2515-683X
 - 5 Hochschule Rhein-Waal; nicki.marquardt@hsrw.eu; ORCID 0000-0001-6501-2692

Behaviorally Anchored Rating Scales (BARS) for Assessing Student Collaboration

Abstract

Universities have the responsibility to effectively prepare students to solve complex problems and challenges in transdisciplinary teams. This paper discusses the potential of Behaviorally Anchored Rating Scales (BARS) for measuring and promoting collaboration in higher education. Through an iterative process, ten behaviorally anchored scales were developed. These scales were evaluated with the involvement of various expert groups (researchers, teachers, students). The results suggest that the behavioral anchors are generally representative and easy to understand. Additionally, potential applications of BARS in the higher education context are outlined.

Keywords

collaboration, higher education, behaviorally anchored rating scales (BARS), instrument

1 Studentische Kollaboration im Hochschulkontext

Kollaboration gilt als eine der Schlüsselkompetenzen für das Leben und Arbeiten der Zukunft. Herausforderungen und Probleme werden in enger Zusammenarbeit, sowie durch den Austausch und das Weiterdenken von Ideen gelöst. Dementsprechend ist die Fähigkeit zu kollaborieren auch bei Arbeitgeber:innen gefragt (Rios et al., 2020). Allerdings stellen diese oft fest, dass es eine Diskrepanz zwischen ihrer Erwartung und den tatsächlichen Kompetenzen von Hochschulabsolvent:innen gibt (sog. *Soft Skills Gap*; Singh Dubey et al., 2022). Potenzielle Gründe hierfür liegen u. a. darin, dass Studierende und Hochschulabsolvent:innen einerseits dazu neigen, die Relevanz von Schlüsselkompetenzen wie Kollaboration zu unterschätzen (Succi & Canovi, 2020) und sich andererseits durch ihre Hochschulen bzw. die Umsetzung der Curricula nicht gut ausgebildet fühlen (z. B. Le et al., 2018).

Kollaboration umfasst kognitive, regulative, affektive sowie behaviorale Prozesse und zeichnet sich durch die wechselseitige Interaktion mehrerer Personen aus (Schürmann et al., 2024). Kollaboration ist abzugrenzen von Kooperation, bei welcher der Fokus auf der Aufteilung der Arbeit liegt und Studierende parallel an Teilaufgaben arbeiten, um diese schließlich zusammenzufügen (Dillenbourg, 1999).

Gruppenarbeiten sind im Hochschulkontext ein verbreitetes Mittel, um Kollaboration zu fördern, indem sie den Austausch und die gemeinsame Problemlösung anregen. Jedoch kommt die Frage auf, ob die Studierenden tatsächlich miteinander kollaborieren oder eher in einem kooperativen Rahmen agieren. Ferner zeigen sowohl Forschung als auch Praxis, dass Studierende dabei häufig auf sich allein gestellt sind und nur wenige Lehrpersonen mehr als die Gruppeneinteilung zu Beginn begleiten (auch als *Sink or Swim*-Ansatz bezeichnet; Vik, 2001; Xu, 2024). Oft besteht seitens der Lehrenden die Erwartungshaltung, dass Studierende soziale Kompetenzen und Kollaborationsfähigkeiten automatisch dadurch erlernen, dass sie in Gruppen eingeteilt werden und über das Semester hinweg zusammenarbeiten (Leopold & Smith, 2020). Der Fokus kollaborativer Lernansätze liegt vor allem auf *Collaborating to*

Learn statt *Learning to Collaborate* (siehe Salmons, 2019). Studierende sowie Lehrende konzentrieren sich bei Gruppenprojekten meist auf inhaltliche statt auf gruppendynamische Aspekte. Dabei gibt es neben kognitiven, insbesondere motivationale, soziale und affektive Herausforderungen für Studierende wie unterschiedliches Interesse und Commitment der Teammitglieder, inkompatible Arbeitsstile und interpersonelle Konflikte (z. B. Näykki et al., 2014). Um Teamprozesse erfolgreich zu gestalten und zu lernen, was eine effektive und gewinnbringende Zusammenarbeit ausmacht, bedarf es folglich sowohl für Studierende als auch für Lehrende mehr Unterstützung beim *Learning to Collaborate*, z. B. in Form von Instruktion, Training und Feedback.

Insbesondere im Hochschulkontext ist dies bisher jedoch nur selten der Fall. Die Lehrenden können eine intensive Betreuung aufgrund großzügig besetzter Kurse oft nicht gewährleisten. Im Vergleich zu Schulen gibt es an Hochschulen außerdem folgende Besonderheiten: Einerseits wird von den Studierenden selbstreguliertes Lernen erwartet. Andererseits sind die Lehrpersonen an Hochschulen in erster Linie fachliche statt didaktische Expert:innen und sind pädagogisch weniger ausgebildet als Lehrkräfte an Schulen, die ein Referendariat durchlaufen (siehe auch Riebe et al., 2016).

Instruktionen darüber, was eine erfolgreiche Zusammenarbeit ausmacht sowie die Integration von Feedback sind vielversprechende Ansätze zur Förderung von Kollaboration (z. B. Eshuis et al., 2019). Auch die Nutzung unterstützender Tools wie Group Awareness Tools stellt eine wertvolle Möglichkeit dar, kollaborative Lern- und Arbeitsprozesse zu verbessern. Diese Tools erfassen den Status quo und bilden diesen transparent ab, sodass das Bewusstsein über die Aktivitäten, Rollen und Fortschritte der Gruppenmitglieder erhöht wird (Bodemer & Dehler, 2011). Um Kollaborationsprozesse und -fähigkeiten gezielt zu fördern, ist es sinnvoll, diese zunächst zu messen und darauf aufbauend Verbesserungspotenziale abzuleiten.

Im Hochschulkontext dauern Gruppenprojekte oft mehrere Wochen an (Schürmann et al., 2024) und enden mit einer gemeinsamen Prüfungsleistung. Der Fokus im Rahmen solcher Projekte liegt in der Regel auf inhaltlichen Aspekten sowie fachlichen

und methodischen Kompetenzen, da sich die Messung kollaborativer Prozesse und Fähigkeiten für Lehrende als schwierig gestaltet (Le et al., 2018). Dies lässt sich u. a. durch die Komplexität des Konstrukts Kollaboration erklären sowie durch die Tatsache, dass für Lehrende meist nur Teilaspekte der studentischen Zusammenarbeit sichtbar sind (z. B. aufgrund von Gruppentreffen außerhalb der Kurszeiten). Mithilfe geeigneter psychologischer Messinstrumente, können sowohl Lehrende als auch die Studierenden selbst, die kollaborative Lern- und Kompetenzentwicklung besser im Blick haben, diese reflektieren und Entwicklungspotenziale erkennen. Die vorliegende Studie hebt hervor, inwiefern verhaltensverankerte Einstufungsskalen studentische Kollaboration gewinnbringend begleiten können. Ferner wird die Entwicklung und Evaluation eines solchen Instruments beschrieben. Ziel ist es, eine fundierte Grundlage für die Messung und Förderung kollaborativer Lernprozesse zu schaffen und dadurch die Qualität der Zusammenarbeit in Gruppenprojekten nachhaltig zu verbessern.

2 Verhaltensverankerte Einstufungsskalen zur Erfassung und Förderung studentischer Kollaboration

Verhaltensverankerte Einstufungsskalen (engl. *Behaviorally Anchored Rating Scales*, BARS) sind ein in der Psychologie etabliertes Bewertungsformat für die Leistungsbeurteilung und Verhaltensevaluation. BARS wurden von Smith und Kendall (1963) eingeführt, um Beobachtungs- und Bewertungsverfahren zu verbessern und zu standardisieren. BARS bestehen aus Skalen mit Verhaltensankern als Antwortoptionen, die verschiedene Qualitätsstufen eines Konstrukts darstellen, wobei hohe Werte für effektive Verhaltensweisen stehen (Debnath et al., 2015). In der Regel sind die Skalen dabei fünf- bis zehnstufig ausgeprägt. BARS finden in verschiedenen Kontexten Anwendung (Debnath et al., 2015), wurden aber vor allem für die Bewertung von Kompetenzen und Leistung im organisationalen Kontext entwickelt und

eingesetzt (Bernardin et al., 1976). Forschung zur Erprobung von BARS im Hochschulkontext ist rar, aber vielversprechend (McIntyre & Gilbert, 1994; Ohland et al., 2012). Beispielsweise schätzten Studierende die in einem Managementkurs zur Bewertung und Benotung eingesetzten BARS als sehr hilfreich und fair ein (McIntyre & Gilbert, 1994).

BARS gelten gemeinhin als objektiver als die meisten Beobachtungsskalen (Anshel, 2016). Forschungsergebnisse deuten darauf hin, dass auf BARS basierendes Feedback im Vergleich zu alternativen Bewertungsmethoden größere Verhaltensänderungen bei den Beurteilten bewirkt (Hom et al., 1982). Während einige Studien die Reliabilität und Validität von BARS belegen (Harrell & Wright, 1990), bleibt jedoch unklar, ob BARS alternativen Bewertungsinstrumenten in der psychometrischen Güte überlegen sind (Hom et al., 1982).

BARS erscheinen für die Messung und Reflexion kollaborativer Prozesse besonders geeignet, da sie es ermöglichen, Konstrukte zu messen, die nicht direkt beobachtbar (Anshel, 2016) und multidimensional sind (Landy et al., 1976, in Georganta & Brodbeck, 2020). Zudem beziehen sie sich auf die Qualität von Verhaltensweisen (Georganta & Brodbeck, 2020), was wiederum ermöglicht, Verbesserungs- und Wachstumspotenziale zu identifizieren oder positive Verhaltensweisen und Leistungen anzuerkennen (Anshel, 2016). In Anbetracht der vielversprechenden Ergebnisse zu ihrem Feedback-Potenzial (Hom et al., 1982), ihrem Einsatz für Selbst- und Peer-Evaluationszwecke (Ohland et al., 2012) sowie in Gruppenkontexten (Georganta & Brodbeck, 2020), sind wir überzeugt davon, dass BARS im Hochschulkontext vielseitig und gewinnbringend eingesetzt werden können.

3 Methodisches Vorgehen

In der Forschungsliteratur finden sich verschiedene Verfahren und Formate für die Entwicklung von BARS (Bernardin et al., 1976), was auf die Flexibilität des Entwicklungsprozesses hindeutet (Debnath et al., 2015). Da BARS häufig zur Leistungsbeurteilung im Arbeitskontext eingesetzt werden, beginnt der Konstruktionsprozess in der Regel mit den Feststellungen der Anforderungen und erfolgskritischer Verhaltensweisen über die Critical Incident Technique (Kanning, 2018). Hierzu werden Fachexpert:innen, auch Subject Matter Experts (SMEs) genannt, die mit der jeweiligen Stelle vertraut sind, interviewt.

Der in der vorliegenden Studie genutzte Ansatz zur Erstellung der BARS ist angelehnt an Georganta und Brodbeck (2020), die verhaltensverankerte Skalen für einen ähnlichen Anwendungsfall (Evaluation von Teamanpassungsprozessen) entwickelt haben. Es wurde ein deduktiver Ansatz verfolgt, der sich aus den folgenden Schritten zusammensetzt: Um eine theoretische Grundlage zu schaffen wurde Kollaboration zunächst auf Basis einer Literaturanalyse definiert und konzeptualisiert. Das daraus entstandene Kodierschema wurde in Pilotstudien getestet und zur Herleitung und Klassifikation von Verhaltensankern genutzt. Diese Initialversion der BARS wurde anschließend durch eine Fragebogenkonferenz überarbeitet. Zur abschließenden Prüfung erfolgte eine Evaluation und Analyse des Instruments mit Hilfe von SMEs.

3.1 Erstellung der BARS

Auf der Grundlage einer systematischen Literaturanalyse wurden zehn Facetten von Kollaboration abgeleitet, die sich vier Kategorien zuordnen lassen (Schürmann et al., 2024). Abbildung 1 bietet eine komprimierte Übersicht dieser Facetten, die die zentralen Aspekte visualisiert und zusammenfasst, ohne deren vollständige Komplexität abzubilden. Im Rahmen darauffolgender Pilotstudien wurde die Anwendbarkeit und Prävalenz dieser Facetten geprüft, indem sie als Kodierschema für die Beobachtung studentischer Kollaboration genutzt wurden. In der ersten Pilotstudie wurden drei Studierendengruppen beobachtet. Es wurden Videoaufnahmen von Gruppen mit ca.

sechs Psychologiestudierenden genutzt, die gemeinsam mehrere Wochen an einem empirischen Forschungsprojekt gearbeitet haben. Durch die Beobachtung wurde die Anwendbarkeit und Passung der Facetten bestätigt. Zudem konnten dadurch die Beschreibungen der jeweiligen Facetten optimiert werden.

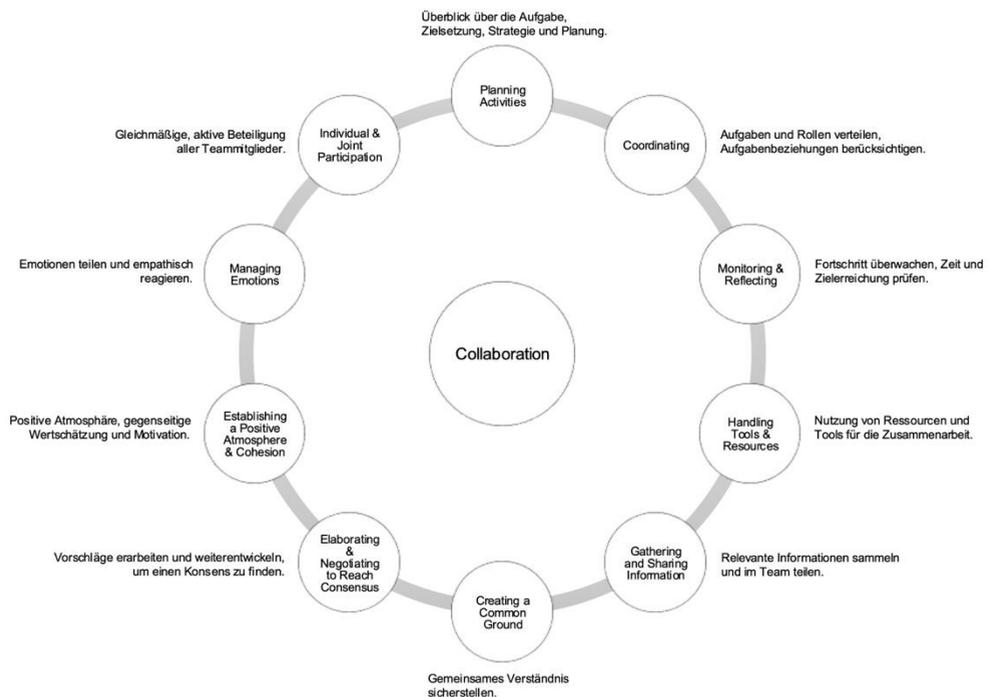


Abb. 1: Zusammenfassende Übersicht der Facetten studentischer Kollaboration (basierend auf Schürmann et al., 2024)

In einer weiteren Pilotstudie wurden Critical Incident Technique (Flanagan, 1954) und Häufigkeitsanalysen kombiniert, um Verhaltensanker für jede der zehn Facetten abzuleiten und zu klassifizieren (Schürmann et al., 2023). Dazu wurden die Video-

aufnahmen von zwei Studierendengruppen verwendet, die über mehrere Stunden virtuell an einer kollaborativen Problemlösungsaufgabe arbeiteten. Zudem schätzten sie die Qualität ihrer Kollaboration anhand eines Fragebogens basierend auf dem Framework von Schürmann et al. (2024) selbst ein. Die Auswahl der zwei Gruppen aus einem größeren Datensatz basierte auf einem *Contrasting Cases*-Ansatz: Eine der Gruppen hatte die höchsten Werte in der Selbstbewertung ihrer Kollaboration, die andere Gruppe die niedrigsten Werte. Auf diese Weise konnten Verhaltensweisen von mangelhaft bis ausgezeichnet identifiziert werden. Durch die Kombination qualitativer und quantitativer Erkenntnisse entstanden zehn verhaltensverankerte Skalen, eine für jede der zehn Facetten von Kollaboration. Dabei wurden 5-stufige Skalen mit Anker für niedrige, mittlere und hohe Ausprägungen gewählt. Die Bewertungen „1“, „3“ und „5“ umfassten jeweils vier Verhaltensanker, die mangelnde, befriedigende bzw. ausgezeichnete Verhaltensweisen widerspiegeln. Die Stufen „2“ und „4“ enthielten keine Verhaltensanker, sondern können gewählt werden, wenn ein Verhalten gezeigt wird, das zu den beiden angrenzenden Stufen passt. Dieses Design ist typisch und hat sich in früheren Untersuchungen als geeignet für BARS erwiesen (z. B. Georganta & Brodbeck, 2020; Ohland et al., 2012).

3.2 Evaluation der BARS

Um die initiale Version der BARS zu überprüfen und zu optimieren, wurde eine Fragebogenkonferenz sowie eine Retranslationsstudie durchgeführt. Im Sinne partizipativer Forschung wurden insbesondere die Perspektiven Studierender mit einbezogen. Im Rahmen der Fragebogenkonferenz wurde zunächst die Passung und Verständlichkeit der BARS diskutiert und Überarbeitungen vorgenommen. Im nächsten Schritt wurden Subject Matter Experts (SMEs, siehe 3.2.1) zu Rate gezogen und gebeten, eine Retranslation (d. h. Zuordnung) der Verhaltensbeispiele zu den passenden Dimensionen und Effektivitätsstufen vorzunehmen (z. B. Smith & Kendall, 1963; Georganta & Brodbeck, 2020). Dies geschah mit Hilfe einer Onlinebefragung.

3.2.1 Stichprobe

An der Fragebogenkonferenz nahmen vier Studierende (zwei Projektbeteiligte, zwei Unbeteiligte) teil. In der darauffolgenden Onlinebefragung wurden drei Untergruppen an SMEs einbezogen, um die verschiedenen Perspektiven potenzieller Anwendergruppen (Forschende im Bereich Teamarbeit/ Kollaboration, Lehrende an Hochschulen, Studierende) zu berücksichtigen. Hierbei konnten insgesamt $N = 24$ Studiendurchgänge in die Analyse aufgenommen werden. Die Altersspanne dieser Teilnehmenden lag zwischen 21 und 51 Jahren ($M = 32.8$, $SD = 11$). Die Stichprobe bestand zu 54 % aus Frauen und zu 46 % aus Männern. Es nahmen insgesamt fünf Forschende teil, die im Durchschnitt sieben Jahre Forschungserfahrung aufwiesen ($SD = 6.9$). Die Gruppe der Lehrenden im Hochschulkontext setzte sich aus sieben Teilnehmenden mit einer durchschnittlichen Lehrerfahrung von 14,3 Jahren ($SD = 4.8$) zusammen. Zudem nahmen 12 Studierende teil, die durchschnittlich 7,3 Hochschulsemeister absolviert hatten ($SD = 4.2$).

3.2.2 Ablauf

Die Fragebogenkonferenz fand in einem virtuellen Konferenzraum statt. Die vier Beteiligten analysierten die BARS Schritt für Schritt, um etwaige Unstimmigkeiten festzustellen und zu korrigieren. Die in diesem Prozess vorgebrachten Verbesserungsvorschläge dienten als Grundlage für die Überarbeitung des Fragebogens (Döring & Bortz, 2016).

Innerhalb der Retranslation bestand die Aufgabe für die Teilnehmenden darin, die einzelnen Verhaltensbeispiele mit Hilfe der Dimensionsdefinitionen der dazugehörigen Dimension der BARS zuzuordnen. Die Reihenfolge der 93 Verhaltensbeispiele der BARS wurde im Vorhinein randomisiert. Zudem erfolgte eine Effektivitätsbewertung der Verhaltensbeispiele. Hierbei wurden die Teilnehmenden aufgefordert, das jeweilige Verhaltensbeispiel der Ausprägung „ausgezeichnet“, „befriedigend“ oder „mangelhaft“ in Bezug zu der ausgewählten Dimension zuzuordnen. Die Teilnehmenden wurden darauf hingewiesen, dass sie etwa 60–90 Minuten benötigen würden, während die tatsächliche Bearbeitungszeit im Durchschnitt unter einer Stunde lag.

3.2.3 Auswertung

Die Retranslation wurde anschließend mittels Häufigkeitsanalyse und einer Berechnung der prozentualen Übereinstimmung der Bewertenden für die einzelnen Verhaltensbeispiele ausgewertet (z. B. Smith & Kendall, 1963; Wright, 1986). Bei der Berechnung der prozentualen Übereinstimmung wird geprüft, wie hoch der Anteil an Bewertenden ist, die die angedachte Dimension zu dem einzelnen Verhaltensbeispiel zugeordnet haben. Innerhalb der bisherigen Forschung existieren verschiedene Grenzen, ab welcher prozentualen Übereinstimmung der Bewertenden eine Eliminierung des Verhaltensbeispiels aus dem finalen Instrument in Betracht gezogen werden sollte. Diese Grenzen bewegen sich in den meisten Studien zwischen 80 % Übereinstimmung (Grussing et al., 1994) und der einfachen Mehrheit (z. B. Klieger et al., 2018; Smith & Kendall, 1963). Hohe Reliabilitätsstandards sind vor allem dann entscheidend, wenn wichtige Konsequenzen (bspw. Entscheidungen innerhalb eines Gerichtsverfahrens) aufgrund des Messergebnisses in Betracht gezogen werden (Krippendorff, 2018). Da die Nutzung der vorliegenden BARS nicht intendiert, lebensverändernde Entscheidungen auf Grundlage der Messergebnisse zu treffen, wurde in der vorliegenden Evaluation die einfache Mehrheit als Grenzwert festgelegt. Die gleiche Auswertung findet zudem bei der Effektivitätsbewertung statt. Zwei fehlende Werte wurden durch die Ermittlung des Medians als Schätzwert ersetzt.

4 Ergebnisse

Die Diskussion über Passung und Verständlichkeit der BARS durch die Fragebogenkonferenz zeigte einerseits, dass die initiale Version u. a. als „zu detailliert“ und zum Teil als „schwierig zu beantworten“ bewertet wurde. Andererseits wurde der Nutzen der detaillierten Dimensionsbeschreibungen für das Verständnis von Kollaboration betont. Die Verhaltensbeispiele wurden positiv bewertet, da sie den Studierenden helfen können, ihr Verhalten innerhalb von Gruppenarbeiten besser einzuordnen und anzupassen. Es wurden Vorschläge zur Änderung einzelner Verhaltensbeispiele eingebracht, die zur Überarbeitung der BARS genutzt wurden. Insgesamt wurden auf Basis der Fragebogenkonferenz 27 Verhaltensbeispiele entfernt oder

kombiniert, wodurch die Gesamtanzahl von 120 auf 93 reduziert wurde. Abbildung 2 skizziert den Aufbau und stellt einen beispielhaften Auszug der BARS dar. Jede Skala besteht aus einer kurzen Beschreibung der Facette und einer 5-stufigen Bewertungsskala mit Verhaltensankern, die verschiedene Qualitätsausprägungen widerspiegeln.

Die vollständige, überarbeitete Version der BARS ist dem Artikel als Zusatzmaterial beigelegt.

Establishing a Positive Atmosphere	
In dem Team herrscht eine positive Atmosphäre. Es wird höflich und freundlich kommuniziert, das heißt, es wird sich begrüßt, bedankt und bei Unterbrechungen entschuldigt. Die Teammitglieder hören sich aktiv zu und schenken sich gegenseitig Anerkennung und Wertschätzung. Es herrscht ein motivierendes Wir-Gefühl („Wir schaffen das“).	
5	<ul style="list-style-type: none"> • Das Team geht respektvoll miteinander um und formuliert kritische Anmerkungen konstruktiv und sachlich. • Das Team ermutigt sich gegenseitig, seine Meinungen und Sichtweisen einzubringen (bspw. durch das Loben der Ideen von Mitgliedern). • Innerhalb des Teams wird die Einstellung „Gemeinsam schaffen wir das“ geteilt.
4	<ul style="list-style-type: none"> • Das Team zeigt Verhaltensweisen, die sowohl bei 3 als auch bei 5 beschrieben werden.
3	<ul style="list-style-type: none"> • Das Team geht gelegentlich respektvoll miteinander um und formuliert kritische Anmerkungen in manchen Fällen konstruktiv und sachlich. • Das Team ermutigt sich manchmal gegenseitig, seine Meinungen und Sichtweisen einzubringen (bspw. durch das Loben der Ideen von Mitgliedern). • Innerhalb des Teams wird die Einstellung „Gemeinsam schaffen wir das“ teilweise geteilt.
2	<ul style="list-style-type: none"> • Das Team zeigt Verhaltensweisen, die sowohl bei 1 als auch bei 3 beschrieben werden.
1	<ul style="list-style-type: none"> • Das Team geht nicht respektvoll miteinander um und formuliert kritische Anmerkungen nicht konstruktiv und sachlich. • Das Team ermutigt sich nicht gegenseitig, seine Meinungen und Sichtweisen einzubringen (bspw. kein Lob für die Ideen von Mitgliedern). • Innerhalb des Teams wird die Einstellung „Gemeinsam schaffen wir das“ nicht geteilt.

Abb. 2: Beispielhafte Darstellung einer verhaltensverankerten Skala (BARS)

Die Analyse der Retranslation zeigt innerhalb der Gesamtstichprobe, dass 78 von 93 Verhaltensankern von den Bewertenden zu der angedachten Dimension zugeordnet wurden. Bei 15 Verhaltensbeispielen befindet sich die prozentuale Übereinstimmung der Bewertenden unter 50 %. Bei der Dimension *Elaborating and negotiating to reach consensus* herrschte am wenigsten Übereinstimmung. Innerhalb der Effektivitätsbewertung wählten über 50 % der Bewertenden bei 91 von 93 Verhaltensbeispielen die angedachte Ausprägung aus. Zwei Verhaltensbeispiele der befriedigenden Ausprägung wurden nicht von der Mehrheit korrekt zugeordnet. Die Verhaltensanker mit unzureichender Übereinstimmung sind im beigefügten Zusatzmaterial, der vollständigen BARS-Version, durch Fußnoten hervorgehoben.

Die Ergebnisse der Retranslation können zur weiteren Überarbeitung der Skala genutzt werden. Die Verhaltensbeispiele mit einer (zu) geringen prozentualen Übereinstimmung der Bewertenden sollten demnach genauer betrachtet und überarbeitet oder aus der Skala entfernt werden.

5 Diskussion

Hochschulen stehen in der Verantwortung, Studierende effektiv darauf vorzubereiten, komplexe Probleme und Herausforderungen in interdisziplinären Teams zu lösen. Der Beitrag beleuchtet das Potenzial von verhaltensverankerten Einstufungsskalen (BARS) zur Messung und Förderung von Kollaboration im Hochschulkontext und stellt entsprechende Skalen vor. Die Evaluation der Skalen zeigt, dass die meisten Verhaltensanker für potenzielle Anwender:innen nachvollziehbar sind. Einige Anker sollten jedoch kritisch hinterfragt werden, insbesondere in der Dimension *Elaborating and negotiating to reach consensus*.

Die in der vorliegenden Studie genutzte Kombination aus literaturgestütztem und empirischem Vorgehen bietet eine solide Grundlage für die Entwicklung von BARS. Während die Literaturanalyse eine fundierte theoretische Basis sicherstellt, half die Critical Incident Technique praxisnahe Verhaltensanker zu identifizieren. Eine wei-

tere Stärke der Studie stellt der iterative Entwicklungsprozess dar, bei dem die initiale Version der Skalen systematisch durch Fragebogenkonferenz und Retranslation überprüft wurde. Dieser mehrstufige Ansatz förderte die kontinuierliche Optimierung der Skalen und führte zur Kürzung sowie einer besseren Verständlichkeit. Die Einbeziehung unterschiedlicher Expertengruppen (Forschende, Lehrende, Studierende) berücksichtigt die Perspektiven potenzieller Anwender:innen der Skalen, was die Akzeptanz und Relevanz der Skalen im praktischen Einsatz stärkt. Die Ergebnisse weisen darauf hin, dass das Instrument die Erfassung von Kollaboration repräsentativ abbilden könnte und liefern erste Hinweise auf eine potenzielle Inhaltsvalidität.

5.1 Limitationen der Studie

Dass die gleiche Stichprobe für die Retranslation und die Effektivitätsbewertung der Verhaltensbeispiele genutzt wurde, kann kritisch betrachtet werden, da Smith und Kendall (1963) zwei voneinander unabhängige Stichproben hierfür empfehlen. In der vorliegenden Studie wurde dieser Schritt aus ökonomischen Gründen kombiniert. Auch wenn dies in der Forschungsliteratur gängig ist (z. B. Grussing et al., 1994; Wright, 1986), ist eine separate Bewertung vermutlich vorzuziehen. So merkten einzelne Teilnehmende an, dass sie die Studienbearbeitung als anstrengend und lang empfanden. Um die Antwortqualität sicherzustellen und negative Erfahrungen von Teilnehmenden zu verhindern, ist daher eine Anpassung des Designs zu empfehlen. Trotz eines vorherigen Pretests zeigte sich ferner, dass die Aufgabenstellung zur Effektivitätsbewertung zum Teil missverstanden wurde. Dies äußerte sich durch aufkommende Fragen einer Teilnehmerin. Es kann nicht ausgeschlossen werden, dass weitere SMEs die Aufgabenstellung fehlinterpretiert haben. Auch wenn die Durchführung der Studie mit Hilfe eines Online-Fragebogens durchaus ökonomische Vorteile liefert, ist für zukünftige Designs ein Präsenzformat mit stärkerer Anleitung und Unterstützung zu empfehlen.

5.2 Implikationen

Das in dieser Studie entwickelte und evaluierte Instrument kann vielseitig und gewinnbringend im Hochschulkontext eingesetzt werden. Die Skalen sind z. B. für Lehrende von Interesse, um kollaborative Lern- und Arbeitsprozesse von Studierenden gezielt zu beobachten und systematisch zu evaluieren. Dies könnte bspw. durch die strukturierte Beobachtung von Gruppenarbeiten in Echtzeit oder anhand von Aufzeichnungen geschehen. Diese Bewertungen wiederum können als Grundlage für Feedbackgespräche mit Studierenden dienen und zur Identifikation von Verbesserungspotenzialen genutzt werden. Die spezifischen Verhaltensanker der BARS werden dabei sowohl von Feedbackgebenden als auch den Empfänger:innen geschätzt (MacDonald & Sulsky, 2009). In Anbetracht hoher Kursgrößen und begrenzter Ressourcen der Lehrenden ist dieser Ansatz ggf. nur begrenzt möglich. In Erweiterung dessen schlagen wir daher vor, dass die vorgestellten BARS von Studiengruppen zur gemeinsamen Reflexion über ihre Zusammenarbeit genutzt werden können. Durch die Verhaltensanker bieten BARS studentischen Teams die Möglichkeit, einen gemeinsamen Erwartungshorizont und somit einen geteilten Bezugsrahmen zu schaffen (Smith & Kendall, 1963).

Auf diese Weise ermöglichen die BARS, ein gemeinsames Verständnis von kollaborativer Zusammenarbeit zu entwickeln. Durch die Nutzung und Auseinandersetzung mit den BARS lernen die Studierenden, welche Faktoren eine effiziente und eine ineffiziente studentische Zusammenarbeit beeinflussen (Ohland et al., 2012). Da Gruppenarbeiten oft auch außerhalb von Kurszeiten stattfinden, haben die Studierenden selbst vermutlich sogar einen besseren Überblick und tiefere Einblicke in den gesamten Kollaborationsprozess als die Lehrenden. Hierbei gilt es jedoch zu berücksichtigen, dass Studierende mit dem BARS-Format oftmals nicht vertraut sind (Ohland et al., 2012), was eine gewisse Erklärung und Instruktion im Vorfeld erfordert. Eine einfache und praktikable Lösung hierfür könnte darin bestehen, dass ein Mitglied pro Team instruiert bzw. geschult wird, welches dann die Erklärungen innerhalb der Gruppe weitergibt.

Die Selbsteinschätzungen der Studierenden könnten z. B. mithilfe von Awareness Tools an die Lehrenden übermittelt werden, die dadurch ein Bewusstsein über die kollaborativen Lern- und Arbeitsprozesse einzelner Gruppen erhalten. Idealerweise kann so bereits während des Semesters erkannt werden, welche Teams gruppenspezifischen Herausforderungen gegenüberstehen. Die Unterteilung in verschiedene Dimensionen ermöglicht darüber hinaus eine bedarfsorientierte und zielgenaue Unterstützung.

Ein vielversprechender Ansatz für zukünftige Forschung besteht darin, die Praktikabilität und Nützlichkeit von BARS im Hochschulkontext zu untersuchen. Dabei sollten verschiedene Perspektiven, wie die der Lehrenden und der Studierenden, einbezogen werden, um ein umfassendes Bild der Akzeptanz und Anwendbarkeit des Instruments zu erhalten.

Literaturverzeichnis

- Anshel, M. H. (2016). Development of a Rating Scale for Determining Competence in Basketball Referees: Implications for Sport Psychology. *The Sport Psychologist*, 9(1), 4–28.
- Bernardin, H. J., LaShells, M. B., Smith, P. C., & Alvares, K. M. (1976). Behavioral expectation scales: Effects of developmental procedures and formats. *Journal of Applied Psychology*, 61(1), 75–79.
- Bodemer, D., & Dehler, J. (2011). Group awareness in CSCL environments. *Computers in Human Behavior*, 27(3), 1043–1045.
- Debnath, S. C., Lee, B. B., & Tandon, S. (2015). Fifty years and going strong: What makes behaviorally anchored rating scales so perennial as an appraisal method? *International Journal of Business and Social Science*, 6(2), 16–25.
- Dillenbourg, P. (1999). Introduction: What do you mean by “collaborative learning”? In P. Dillenbourg (Ed.), *Collaborative learning: Cognitive and computational approaches* (Vol. 1, pp. 1–19). Elsevier.
- Flanagan, J. C. (1954). The Critical Incident Technique. *Psychological Bulletin*, 51(4).
- Döring, N., & Bortz, J. (2016). Datenerhebung. In N. Döring & J. Bortz (Hrsg.), *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5th ed., pp. 321–578). Springer.
- Eshuis, E. H., ter Vrugte, J., Anjewierden, A., Bollen, L., Sikken, J., & de Jong, T. (2019). Improving the quality of vocational students’ collaboration and knowledge acquisition through instruction and joint reflection. *International Journal of Computer-Supported Collaborative Learning*, 14(1), 53–76.
- Harrell, A., & Wright, A. (1990). Empirical Evidence on the Validity and Reliability of Behaviorally Anchored Rating Scales for Auditors. *Auditing: A Journal of Practice & Theory*, 9(3), 134–149.
- Georganta, E., & Brodbeck, F. C. (2020). Capturing the Four-Phase Team Adaptation Process with Behaviorally Anchored Rating Scales (BARS). *European Journal of Psychological Assessment*, 36(2), 336–347.
- Grussing, P. G., Valuck, R. J., & Williams, R. G. (1994). Development and Validation of Behaviorally-Anchored Rating Scales for Student Evaluation of Pharmacy Instruction. *American Journal of Pharmaceutical Education*, 58(1), 25–37.

- Hom, P. W., DeNisi, A. S., Kinicki, A. J., & Bannister, B. D. (1982). Effectiveness of performance feedback from behaviorally anchored rating scales. *Journal of Applied Psychology, 67*(5), 568–576.
- Kanning, U. P. (2018). *Diagnostik für Führungspositionen*. Hogrefe.
- Klieger, D. M., Kell, H. J., Rikoon, S., Burkander, K. N., Bochenek, J. L., & Shore, J. R. (2018). Development of the Behaviorally Anchored Rating Scales for the Skills Demonstration and Progression Guide. *ETS Research Report Series, 2018*(1), 1–36. <https://doi.org/10.1002/ets2.12210>
- Krippendorff, K. (2018). *Content Analysis: An Introduction to Its Methodology* (4th ed.). SAGE Publications.
- Le, H., Janssen, J., & Wubbels, T. (2018). Collaborative learning practices: Teacher and student perceived obstacles to effective student collaboration. *Cambridge Journal of Education, 48*(1), 103–122.
- Leopold, H., & Smith, A. (2020). Implementing reflective group work activities in a large chemistry lab to support collaborative learning. *Education Sciences, 10*(7), 1–19.
- MacDonald, H. A., & Sulsky, L. M. (2009). Rating formats and rater training redux: A context-specific approach for enhancing the effectiveness of performance management. *Canadian Journal of Behavioural Science, 41*(4), 227–240.
- McIntyre, F. S., & Gilbert, F. W. (1994). Improving Performance in Case Courses: An Argument for Behaviorally Anchored Rating Scales. *Marketing Education Review, 4*(1), 51–58.
- Näykki, P., Järvelä, S., Kirschner, P. A., & Järvenoja, H. (2014). Socio-emotional conflict in collaborative learning—A process-oriented case study in a higher education context. *International Journal of Educational Research, 68*, 1–14.
- Ohland, M. W., Loughry, M. L., Woehr, D. J., Bullard, L. G., Felder, R. M., Finelli, C. J., Layton, R. A., Pomeranz, H. R., & Schmucker, D. G. (2012). The comprehensive assessment of team member effectiveness: Development of a behaviorally anchored rating scale for self- and peer evaluation. *Academy of Management Learning and Education, 11*(4), 609–630.
- Riebe, L., Girardi, A., & Whitsed, C. (2016). A systematic literature review of teamwork pedagogy in higher education. *Small Group Research, 47*(6), 619–664.

- Rios, J. A., Ling, G., Pugh, R., Becker, D., & Bacall, A. (2020). Identifying critical 21st-century skills for workplace success: A content analysis of job advertisements. *Educational Researcher*, 49(2), 80–89.
- Salmons J. (2019). *Learning to collaborate, collaborating to learn. Engaging students in the classroom and online*. Stylus.
- Schürmann, V., Marquardt, N., & Bodemer, D. (2024). Conceptualization and measurement of peer collaboration in higher education: A systematic review. *Small Group Research*, 55(1), 89–138.
- Schürmann, V., Spahn, T., Marquardt, N., & Bodemer, D. (2023, 10. – 15. June). *Evaluating Peer Collaboration in Higher Education: Behaviorally Anchored Rating Scales* [Poster Presentation], ISLS Annual Meeting, Montréal, Canada.
- Singh Dubey, R., Paul, J., & Tewari, V. (2022). The soft skills gap: A bottleneck in the talent supply in emerging economies. *International Journal of Human Resource Management*, 33(13), 2630–2661.
- Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, 47(2), 149–155.
- Succi, C., & Canovi, M. (2020). Soft skills to enhance graduate employability: comparing students and employers' perceptions. *Studies in Higher Education*, 45(9), 1834–1847.
- Vik, G. N. (2001). Doing More to Teach Teamwork Than Telling Students to Sink or Swim. *Business Communication Quarterly*, 64(1), 112–119.
- Wright, A. (1986). Performance Evaluation of Staff Auditors: A Behaviorally Anchored Rating Scale. *Auditing: A Journal of Practice & Theory*, 5(2), 95–108.
- Xu, S. (2024). Groupwork and collaborative learning: Chinese university students' struggles and strategies. *Higher Education Research and Development*, 43(1), 227–242.