Der Einfluss kriterialer, sozialer und individueller Bezugsnormen auf das experimentbezogene Fähigkeitsselbstkonzept

Ramona Atzert¹ · Rebecca John¹ · Angelika Preisfeld¹ · Karsten Damerau¹

Eingegangen: 29. Januar 2020 / Angenommen: 6. Juni 2020 / Online publiziert: 25. Juni 2020
© Der/die Autor(en) 2020

Zusammenfassung

Schüler*innen sind immer weniger von ihren naturwissenschaftlichen Fähigkeiten überzeugt. Um diesem Problem im Kontext naturwissenschaftlichen Unterrichts begegnen zu können, gilt es, die auf das naturwissenschaftliche Fähigkeitsselbstkonzept (FSK) wirkenden Faktoren zu identifizieren. Individuelle, soziale und kriteriale Bezugsnormen üben einen nachweisbaren Einfluss auf schulische FSK aus. Welche Bezugsnorm Schüler*innen zur Einschätzung ihres experimentbezogenen FSK heranziehen ist Gegenstand der vorliegenden Studie. Diese wurde mit 169 Schüler*innen der gymnasialen Oberstufe realisiert. Hierzu wurde der FbeFSK (Fragebogen zur Erfassung des bezugsnormbelasteten experimentbezogenen FSK) eingesetzt, welcher – in Orientierung an den Skalen zur Erfassung des schulischen Selbstkonzepts (SESSKO) – das experimentbezogene FSK in drei bezugsnormbelasteten Skalen (individuell, sozial, kriterial) und einer absoluten Skala erfasst. Dessen Validität wurde faktoranalytisch überprüft und die Reliabilität optimiert. Die resultierenden Skalen wurden in multiplen Regressionsanalysen eingesetzt, um den Einfluss der drei bezugsnormbelasteten Skalen auf das experimentbezogene FSK zu untersuchen. Ein besonders hoher positiver Einfluss des kriterialbelasteten FSK auf das absolute experimentbezogene FSK kann festgestellt werden. Der Vergleich der standardisierten Regressionskoeffizienten zeigt jedoch kaum einen Einfluss der beiden anderen Prädiktoren. Neben der Qualität der Skalen und den Ursachen der Befunde werden Implikationen für die Praxis des naturwissenschaftlichen Experimentalunterrichts diskutiert.

Schlüsselwörter Bezugsnormen · Bezugsnormorientierung · Fähigkeitsselbstkonzept · Experimentierkompetenz

✉ Karsten Damerau
kdamerau@uni-wuppertal.de

¹ Zoologie & Didaktik der Biologie, Bergische Universität Wuppertal, Wuppertal, Deutschland

The Influence of the Criterial, Social, and Temporal Reference Norm on Students' Experiment-Related Self-Concept

Abstract

Pupils are less and less convinced of their scientific skills. In order to be able to address this problem in science education, it is important to identify the factors that affect the scientific self-concept (FSK). Temporal, social and criterial reference norms have a demonstrable influence on academic self-concepts. The subject of this study is to find out which reference norm pupils use to assess their experiment-related self-concept. This was tested with a sample of 169 pupils from upper secondary schools. For this purpose, a questionnaire which gathered pupils' experimental-related self-concepts influenced by the different reference norms (FbeFSK) was used. The questionnaire, which is based on the Scales for Gathering the Academic Self-Concept (SESSKO), covers four scales—the experiment-related self-concept effected by temporal, social or criterial reference norms, and an absolute scale. Its validity was confirmed through factor analysis and the reliability was optimized. Multiple regression is used to analyse the resulting scales and to investigate the influence of the three reference norms on the experiment-related self-concept. The self-concept influenced by the criterial reference norm has a major positive influence on the experiment-related self-concept. However, the comparison of the standardized coefficients shows hardly any influence of the other two predictors. In addition to reviewing the quality of the scales and naming possible causes, this paper discusses practical consequences for science education.

Keywords Reference norms · Reference norm orientation · Academic self-concept · Experimental competencies

Theoretischer Hintergrund: Bezugsnormen und ihre Relevanz für das experimentbezogene Fähigkeitsselbstkonzept

Schüler und Schülerinnen (Schüler*innen) in Deutschland sind immer weniger von ihren naturwissenschaftlichen Fähigkeiten überzeugt (vgl. Schiepe-Tiska et al. 2016). Diese Entwicklung erscheint bedenklich, da Annahmen über die Höhe der eigenen Fähigkeiten in vielfältiger Weise das Erleben und Verhalten (Meyer 1984), das Interesse (Dennissen et al. 2007; Ertl et al. 2014), die Erfolgserwartung (Priess-Groben und Hyde 2017) und die Leistungsmotivation (Eccles et al. 1983) beeinflussen. So werden auch Kurs-, Berufs- und Karrierewahlen durch Fähigkeitsselbstkonzepte (FSK) bestimmt, da Erfolge prognostiziert werden, die als Entscheidungskriterien dienen (Köller et al. 2000; Ratschinski 2009; Ertl et al. 2014). Einen positiven Zusammenhang zwischen dem Fähigkeitsselbstkonzept und akademischer Leistung bestätigen Meta-Analysen (Hansford und Hattie 1982; Mabe und West 1982; Schöne et al. 2003; Köller et al. 2006). Ein Ziel des naturwissenschaftlichen Unterrichts sollte daher die Herausbildung eines realistisch-positiven naturwissenschaftlichen FSK sein (vgl. Schiepe-Tiska und Schmidtner 2013).

Das Fähigkeitsselbstkonzept beschreibt primär die aus deklarativem Wissen hervorgehende Einschätzung der eigenen Leistungsfähigkeit und weniger die Bewertung oder die Akzeptanz der eigenen Person (Helmke 1992; Spinath 2011). Mit der Einschränkung der Ausblendung der affektiven Komponente des allgemeinen Selbstkonzepts lässt sich das FSK – welches auch als akademisches oder schulisches

Selbstkonzept (Shavelson et al. 1976; Marsh und Shavelson 1985; Stiensmeier-Pelster und Schöne 2008) bezeichnet wird – als multidimensionales, hierarchisches Konstrukt beschreiben, auf dessen unteren Hierarchieebenen sich auch die subjektiven Fähigkeitsüberzeugungen in spezifischen naturwissenschaftlichen Tätigkeiten verorten lassen (vgl. Dickhäuser 2006; Marsh et al. 1988; Shavelson et al. 1976; Autoren Jahreszahl). Aufgrund der zentralen Rolle des Experimentierens im naturwissenschaftlichen Erkenntnisweg (Heinicke und Peters 2014) und im naturwissenschaftlichen Unterricht (Barzel et al. 2012) kommt dem experimentbezogenen FSK im Kontext des naturwissenschaftlichen Unterrichts eine besondere Bedeutung zu. Analog zum Modell experimenteller Kompetenz (Schreiber et al. 2009) beschreibt das experimentbezogene FSK die subjektive Fähigkeitsüberzeugung, Experimente erfolgreich planen, durchführen und auswerten zu können.

Das experimentbezogene FSK von Schüler*innen ist – wie auch andere Facetten akademischer Selbstkonzepte (vgl. Ruploh et al. 2014) – zeitlich nicht stabil. Es kann durch praktischen Experimentalunterricht gesteigert werden (Autoren Jahreszahl). Dass die subjektive Überzeugung von der eigenen Experimentierfähigkeit zwar häufig, aber nicht notwendigerweise immer positiv mit der tatsächlichen experimentellen Performanz korreliert, belegen Schreiber et al. (2016). Bisher bleibt offen, welche Variablen diese Unterschiede in der Selbstbeurteilung der experimentbezogenen Fähigkeiten ausmachen. Relative Referenzrahmen, die von Heckhausen (1974) als Bezugsnormen definiert werden, sind als wesentliche Einflussgrößen von Fähigkeitsselbstkonzepten bekannt (Köller et al. 2006) und daher auch als relevante Einflussfaktoren

experimentbezogener Fähigkeitsselbstkonzepte anzunehmen. Bezugsnormen dienen als Maßstab zur Selbst- und Fremdbewertung von Handlungsergebnissen (Sülz 2014) sowie von Fähigkeitseinschätzungen (Schöne et al. 2003). Eine Differenzierung von individueller, sozialer und kriterialer/sachlicher Bezugsnorm wird im Leistungsbereich vorgenommen (vgl. Heckhausen 1974; Dickhäuser und Rheinberg 2003; Jäger 2008).

Die individuelle Bezugsnorm ist der intrapersonale, temporale Vergleich der aktuellen eigenen akademischen Fähigkeiten zu früheren Fähigkeiten (vgl. Rheinberg 2001; Heckhausen 1974). Aufsteigende oder abfallende Leistungstendenzen im schulischen Kontext begründen demnach motivationale Konsequenzen für das FSK einer Person (Dickhäuser und Rheinberg 2003). Individuelle Vergleiche sollten entsprechend auch für die Ausprägung des experimentbezogenen Fähigkeitsselbstkonzepts Relevanz besitzen. Schüler*innen, welche sich im Laufe ihrer Schullaufbahn als zunehmend fähigere Experimentatoren erleben, sollten über ein höheres experimentbezogenes FSK verfügen, als Schüler*innen mit rückläufiger Experimentierleistung. Die individuelle Bezugsnorm gilt in der Selbstbewertung als vorherrschend bis durch den Eintritt in einen Schulklassenverband ein sozialer Vergleich mit Gleichaltrigen eintritt (vgl. Heckhausen 1974; Brunstein und Heckhausen 2018). Unter der sozialen Bezugsnorm wird der interpersonale Vergleich der eigenen Fähigkeiten mit den Fähigkeiten von Personen des sozialen Umfeldes verstanden (Dickhäuser und Rheinberg 2003). Die entsprechende Bezugsgruppe kann je nach Kontext variieren und in der Schule aus Gleichaltrigen einer Schul- oder Parallelklasse bzw. aus Kursteilnehmer*innen eines gewählten Faches bestehen. Gerade im schulischen Kontext werden die sozialen Vergleiche als wesentlicher Faktor für die Genese des Fähigkeitsselbstkonzepts gesehen (vgl. Helmke 1992; Stiensmeier-Pelster und Schöne 2008). Diese sozialen Bezugsnormen ziehen Personen beispielsweise maßgeblich heran, um den Schwierigkeitsgrad einer zu bewältigenden Aufgabe zu bewerten (Stiensmeier-Pelster und Schöne 2008). Gelingt es der Mehrheit der Bezugsgruppe eine Aufgabe erfolgreich zu lösen, wird diese als einfach bewertet. Die persönliche Einordnung der Schüler*innen hinsichtlich der Aufgabenbewältigung hat nachhaltige Implikationen für ihr FSK (Meyer 1973). Abhängig davon, ob der Vergleich der eigenen Leistungen mit den Leistungen anderer positiv oder negativ ausfällt, wirkt sich dies auch positiv bzw. negativ auf das Konzept der eigenen Fähigkeiten aus (Götz und Preckel 2006). Auf die Relevanz sozialer Vergleiche weist auch der häufig in Unterrichtskontexten zu beobachtende Fischteichereffekt hin. Dieser belegt, dass vergleichbar begabte Schüler*innen in leistungsfähigeren Schulen ein geringer ausgeprägtes FSK entwickeln als in leistungsschwächeren Schulen (vgl. Marsh 1987). Somit ist

anzunehmen, dass Schüler*innen ebenso soziale Vergleiche ihrer Experimentierfähigkeit anstreben, um ihr eigenes experimentbezogenes FSK zu beurteilen. In der Regel wird noch eine weitere, kriteriale (sachliche) Bezugsnorm von den beiden erstgenannten unterschieden. Bei dieser wird die individuelle Leistung mit fixen Richtwerten in Beziehung gesetzt (vgl. Heckhausen 1974; Rheinberg 2001). Leistungen werden demnach bemessen nach den Anforderungen der Sache selbst (Heckhausen 1974). Diese Anforderungen repräsentieren also einen Schwellengrad an notwendiger Bewältigung, um eine Aufgabe erfolgreich zu lösen. Im Kontext des naturwissenschaftlichen Unterrichts können sich diese Anforderungen z. B. in den Zielen und Kompetenzerwartungen der Lehrpläne (vgl. Heckhausen 1974; Dickhäuser und Rheinberg 2003; Stiensmeier-Pelster und Schöne 2008), in den verbalen Aufgabenstellungen durch Lehrkräfte sowie in den schriftlichen Aufgaben auf Arbeitsblättern oder in Experimentieranleitungen manifestieren. Die erfolgreiche Bewältigung kriterialer Anforderungen eines Experimentes sollte dementsprechend in einem höheren experimentbezogenen FSK der Schüler*innen münden.

Das experimentbezogene FSK sollte somit durch die positive Selbstbewertung durch gute Leistungen, die aus individueller Verbesserung (individuelle Bezugsnorm), überdurchschnittlichen Resultaten (soziale Bezugsnorm) oder dem Eintreten eines Lernzieles (kriteriale Bezugsnorm) münden, gesteigert werden (vgl. Dickhäuser und Rheinberg 2003). Bei der Leistungsbeurteilung lassen sich Unterschiede in der Bevorzugung bestimmter Bezugsnormen erkennen. Diese Präferenz wird als Bezugsnormorientierung (BNO) bezeichnet (vgl. Mischo und Rheinberg 1995; Dickhäuser und Rheinberg 2003). Ursprünglich wurden Bezugsnormorientierungen bei der Fremdbewertung (z. B. Lehrer beurteilt Schüler*in) erfasst (Mischo und Rheinberg 1995), sie lassen sich allerdings in gleicher Weise auf die Selbstbeurteilung übertragen (Schöne et al. 2004). Die Bevorzugung einer Bezugsnorm ist abhängig von der beurteilenden Person (Rheinberg et al. 1977). Zudem stellt sie keine ausschließende Dimension dar, da individuelle, soziale oder kriteriale Bezugsnormorientierungen auch gleichzeitig greifen können (Schöne et al. 2004). Empirische Ergebnisse zeigen, dass Grundschulern*innen die kriteriale Bezugsnormorientierung im Zusammenhang mit einer weiteren Bezugsnormorientierung (individuell oder sozial) präferieren, um ihre mathematischen Fähigkeiten zu beurteilen (Lohbeck 2017). Rheinberg et al. (1977) zeigten, dass Schüler*innen mit steigendem Alter – im Vergleich zu sozialen Bezugsnormen – zunehmend individuelle Bezugsnormen heranziehen, um die eigenen Fähigkeiten zu bewerten. Schöne et al. (2004) verorten zusätzlich einen Zusammenhang der Bevorzugung einer Bezugsnorm mit der Art des Zieles einer Handlung, bei der entweder die motivationale Ausrichtung in Kompe-

tenzsteigerung (Lernzielorientierung) oder Kompetenzdemonstration (Leistungszielorientierung) liegt. Grund-, Mittelstufen-, und Oberstufen-Schüler*innen mit einer hohen Lernzielorientierung bevorzugen demnach individuelle Bezugsnormen, leistungszielorientierte Schüler*innen dagegen soziale Bezugsnormen (Schöne et al. 2004). Welche Bezugsnormen Schüler*innen für die Einschätzung ihres experimentbezogenen Fähigkeitsselbstkonzeptes heranziehen soll die im Folgenden berichtete Studie zeigen.

Methode

Ziel der Auswertung war die faktoranalytisch gestützte Generierung von drei bezugsnormbelasteten Skalen (kriterial, sozial, individuell) sowie einer nicht-bezugsnormbelasteten Skala (absolut) zur Erfassung des experimentbezogenen FSK, um anschließend mittels multipler Regression den Einfluss der bezugsnormbelasteten experimentbezogenen FSK auf das absolute experimentbezogene FSK zu untersuchen.

Itementwicklung

Das im Rahmen der vorliegenden Studie von uns neu entwickelte Messinstrument FbeFSK (*Fragebogen zur Erfassung des bezugsnormbelasteten experimentbezogenen FSK*) erhebt das absolute sowie das von sozialen, kriterialen und individuellen Bezugsnormen beeinflusste experimentbezogene FSK in vier Skalen aus jeweils fünfzehn Items. In Orientierung am dreigliedrigen Modell experimenteller Kompetenz nach Schreiber et al. (2009; siehe oben) berücksichtigt jede der vier Skalen sowohl die Fähigkeitsselbsteinschätzung bezogen auf die Planung, Durchführung und Auswertung von Experimenten. Während die Items der Skala *absolut* keinen Vergleichsmaßstab beinhalten (z.B. „Die zu einem Experiment benötigten Geräte kann ich problemlos bedienen.“; „Mit dem Schreiben von Versuchsprotokollen habe ich keine Probleme.“), berücksichtigen die Items der Skala *sozial* den sozialen Vergleich mit Mitschüler*innen (z.B. „Ein Experiment alleine nach einer Fragestellung zu entwickeln fällt mir leichter *als meinen Mitschüler*innen*.“; „Im Bedienen der benötigten Gerätschaften bin ich sicherer *als meine Mitschüler*innen*.“), die Items der Skala *individuell* den Vergleich zu der eigenen früheren Experimentierfähigkeit (z.B. „Grafiken aus Messdaten zu erstellen fällt mir *leichter als früher*.“; „Ein Experiment nach einer vorgegebenen Fragestellung zu entwickeln fällt mir *leichter als früher*.“) und die Items der Skala *kriterial* den Vergleich zu den Ansprüchen des Systems Schule (z.B. „Entstandene Messdaten des Experimentes kann ich so auswerten, dass sie den Anforderungen des Unterrichts gerecht werden.“; „Versuche *fachgerecht* vorzubereiten bereitet mir kei-

ne Schwierigkeiten.“). Während die Inhalte der Items vollständig neu formuliert wurden, orientiert sich die grundlegende Substrukturierung des FbeFSK in eine nicht von Bezugsnormen beeinflusste und in drei bezugsnormbelastete Fähigkeitsselbstkonzept-Skalen an den Skalen zur Erfassung des schulischen Selbstkonzeptes (SESSKO, Schöne et al. 2002, 2012), welche sich als reliables Messinstrument erwiesen haben. Zudem liegen Hinweise auf die faktoriel- le und kriteriale Validität des SESSKO vor (Schöne et al. 2002, 2003, 2012).

Stichprobe und Versuchsdesign

Das zuvor beschriebene Messinstrument wurde zu einem einzelnen Messzeitpunkt bei $N = 169$ Schüler*innen ($N_W = 98 \hat{=} 57,99\%$, $N_M = 71 \hat{=} 42,01\%$) der gymnasialen Oberstufe eingesetzt (Jahrgangsstufen: $N_{10} = 97 \hat{=} 57,40\%$, $N_{11} = 69 \hat{=} 40,83\%$, $N_{12} = 3 \hat{=} 1,78\%$; Durchschnittsalter = 16,3 Jahre). Sämtliche Teilnehmer*innen der Studie hatten mindestens ein naturwissenschaftliches Fach im voraus- gegangenen Schulhalbjahr belegt ($N_{\text{Biologie}} = 140 \hat{=} 82,84\%$, $N_{\text{Chemie}} = 138 \hat{=} 81,66\%$, $N_{\text{Physik}} = 59 \hat{=} 34,91\%$; Mehrfach- nennung möglich). Die Bearbeitungszeit des Fragebogens betrug ca. 15–20 min zuzüglich einiger Minuten für die Instruktion.

Vorgehen bei der Datenauswertung

Vorbereitung des Datensatzes Ein im Anschluss an die Erhebung festgestellter, fehlender kriterialer Bezug bei sechs der im Fragebogendokument vorhandenen fünfzehn Items der Skala kriterial, führte zum Ausschluss dieser sechs Items aus der Auswertung. Die Analyse der Daten erfolgte in SPSS 25. Fehlwerte sowie deren Zufälligkeit (MCAR-Test nach Little, vgl. Little 1988) wurden im Datensatz bestimmt und mittels EM-Algorithmus ersetzt. Ausreißer wurden über die Anomalie-Erkennung in SPSS bestimmt und aus der weiteren Auswertung ausgeschlossen.

Explorative Faktorenanalysen (EFA) inklusive Prüfung ihrer Voraussetzungen Der resultierende Datensatz wurde hinsichtlich seiner Eignung zur Durchführung einer EFA geprüft (Kaiser-Meyer-Olkin-Kriterium, vgl. Kaiser und Rice (1974) und Bartlett-Test, vgl. Bartlett (1951)). Eine konfirmatorische Faktorenanalyse wurde ausgeschlossen, da der neu entwickelte FbeFSK zwar die Substruktur des SESSKO nachempfunden, die vorliegenden Items aber vollständig neu entwickelt wurden. In Orientierung am Vorgehen bei der faktoranalytischen Betrachtung der bezugsnormbelasteten Skalen des SESSKO (vgl. Schöne et al. 2002; Sparfeldt et al. 2003) wurden in der hier vorliegenden Studie zwei explorative Faktorenanalysen (Hauptkomponentenanalyse, Rotation: Varimax) durchgeführt und zwar einmal un-

ter Einbezug der Items des absoluten experimentbezogenen FSK (vier zu extrahierende Faktoren: absolut, kriterial, sozial, individuell) und einmal unter Ausschluss der Items des absoluten experimentbezogenen FSK (drei zu extrahierende Faktoren: kriterial, sozial, individuell). Der dreifaktorielle Ansatz gründet sich in der Überlegung, dass das absolute FSK nicht in dem theoretischen Rahmen der Bezugsnormen zu verorten ist (vgl. Schöne et al. 2002). Da dieser Ausschluss der Bezugsnorm-unabhängigen Items aus einer gemeinsamen Faktorenanalyse aber kritisch gesehen wird (vgl. Sparfeldt et al. 2003), wird in der vorliegenden Studie außerdem der o. g. vierfaktorielle Ansatz realisiert. Für die Einschätzung der tatsächlich vorliegenden Dimensionalität des Messinstrumentes wurde das Elbow-Kriterium herangezogen (vgl. Beavers et al. 2013). Ferner werden die prozentualen Anteile der durch die Faktoren aufgeklärten Gesamtvarianz der Items berichtet. Sowohl die drei- als auch die vierfaktorielle Lösung wurden unabhängig voneinander zur Konstruktion der Skalen sozial, kriterial, individuell und absolut genutzt. Items mit den höchsten Ladungen auf einem gemeinsamen Faktor wurden zu einer Skala zusammengefasst. Lagen die Höchstladungen allerdings auf einem inhaltlich unerwarteten Faktor oder wiesen die Items substantielle Querladungen $\geq 0,4$ auf (vgl. z. B. Noormann 2017), so wurden diese Items aus der weiteren Skalenkonstruktion ausgeschlossen. Faktorladungen unter $< 0,3$ wurden als niedrig gewertet (vgl. Fromm 2012) und werden nicht berichtet. Als Hinweis auf die Stabilität der Faktorenlösungen wurden zudem die Kommunalitäten der Items berechnet. Bei acht oder mehr Items pro Faktor und einer Stichprobengröße von $N = 130$ oder größer werden bereits niedrige (0,20 bis 0,40) bis mittlere Kommunalitäten (0,40 bis 0,60) als hinreichend erachtet, während sich bei Verringerung der Itemzahl pro Faktor die bei gleicher Kommunalität erforderliche Stichprobengröße deutlich erhöht (vgl. Bühner 2011). Da in die EFA mit drei zu extrahierenden Faktoren die Bezugsnorm-unabhängigen Items nicht aufgenommen wurden, wurden im Falle des dreifaktoriellen Ansatzes in die weitere Skalenkonstruktion der Skala *absolut* alle fünfzehn zur Verfügung stehenden Items berücksichtigt.

Prüfung von Trennschärfe und Reliabilität Es folgte eine Überprüfung der Trennschärfen (zufriedenstellend ab $r_{i,i-1}^2 > 0,40$; vgl. Fisseni 1997) und Reliabilitäten (zufriedenstellend ab Cronbachs $\alpha > 0,70$, vgl. Schmitt 1996) zwecks eventueller Optimierung aller Skalen. Die Summenscores der aus dem vier- und dreifaktoriellen Ansatz hervorgehenden Skalen wurden im Rahmen multipler Regressionsanalysen verwendet, um den Einfluss der Bezugsnormbelasteten experimentbezogenen FSK auf das absolute experimentbezogene FSK zu untersuchen.

Multiple Regressionsanalysen inklusive Prüfung ihrer Voraussetzungen Als Voraussetzungen der multiplen Regression wurden die Linearität der Zusammenhänge (Streudiagramme: Bezugsnormbelasteter Prädiktor vs. absolutes experimentbezogenes FSK), Multikollinearität der Prädiktoren (Pearson-Korrelationen und Variance-Inflation-Faktor (VIF), $VIF > 2$ ist tendenziell problematisch, vgl. Schneider 2007) sowie die Ausreißer (Cutoff-Werte für Ausreißer: standardisierte Residuen $> |3|$, vgl. Schlepper 2014; Hebelwerte $> 0,2$, Huber 1981; Cook-Distanz $> 0,5$, vgl. Kohn und Öztürk 2013), Autokorrelationen (Durbin-Watson-Statistik sollte innerhalb 1 und 3 liegen, vgl. Field 2013), Homoskedastizität (Trichterförmige Verteilung im Streudiagrammen – hier standardisierte Residuen vs. vorhergesagte Werte – deutet auf Heteroskedastizität hin, vgl. Tabachnick und Fidell 2014) und Normalverteilung (K-S-Test mit Lilliefors-Korrektur; $p \leq 0,05$ bedeutet eine signifikante Abweichung von der Normalverteilung, vgl. Janssen und Laatz 2017) der Residuen geprüft. Die Signifikanz des Regressionsmodells wird – sowohl mit den Skalen aus der vier- als auch mit den Skalen aus der dreifaktoriellen Lösung – mittels ANOVA bestimmt. Zur Beurteilung der Varianzaufklärung werden jeweils R^2 sowie das korrigierte R^2 berichtet. Die relative Stärke des Einflusses der Bezugsnormbelasteten experimentbezogenen FSK auf das absolute experimentbezogene FSK veranschaulichen die Regressionskoeffizienten.

Ergebnisse

Vorbereitung des Datensatzes Unter Vorliegen der MCAR-Bedingung (MCAR-Test nach Little: $\chi^2(753, N = 169) = 758,36$, n.s.) wurden Fehlwerte ($\leq 1,8\%$ pro Variable) mittels EM-Algorithmus ersetzt. Ausreißer wurden im Anschluss über die Anomalie-Erkennung in SPSS identifiziert. Drei Fälle mit einem – im Vergleich zur übrigen Stichprobe – relativ hohen Einflussmaß der primären Grundvariable ($> 0,09$) beziehungsweise einem relativ hohen Anomalie-Index ($> 1,9$) wurden aus den weiteren Analysen ausgeschlossen. Somit reduzierte sich die Stichprobengröße auf $N = 166$ Schüler*innen.

EFA KMO-Koeffizient ($KMO_4 \text{ Faktor} = 0,87$, $KMO_3 \text{ Faktor} = 0,88$, vgl. Kaiser und Rice 1974) und Barlett-Test ($\chi^2_4 \text{ Faktor}(1431, N = 166) = 5556,96$; $p \leq 0,001$, $\chi^2_3 \text{ Faktor}(741, N = 166) = 3601,07$; $p \leq 0,001$, vgl. Bartlett 1951) bestätigten die Eignung dieses Datensatzes für die Durchführung einer explorativen Faktorenanalyse. Ein Elbow im Verlauf der Eigenwerte deutete bei Einbezug aller Items eine vierdimensionale Struktur des Messinstrumentes an (vgl. Beavers et al. 2013). Tab. 1 zeigt die Faktorladungen und Kommunalitäten der EFA unter Vorgabe vier bzw. drei zu extrahierender Faktoren (vgl. o. g. Methoden, vgl. Sparfeldt et al.

Tab. 1 Faktoren, Ladungen und Kommunalitäten der Skalen zum experimentbezogenen Selbstkonzept

Item	4 Faktoren					3 Faktoren			
	a ₁	a ₂	a ₃	a ₄	<i>h</i> ²	a ₁	a ₂	a ₃	<i>h</i> ²
s1	0,46	0,54	–	–	0,53	0,54	–	0,42	0,48
s2	–	0,68 ^a	–	–	0,56	0,71 ^a	–	–	0,56
s3	–	0,57 ^a	–	0,35	0,50	0,62 ^a	–	–	0,45
s4	0,37	0,66 ^a	–	–	0,64	0,66 ^a	–	0,30	0,56
s5	–	0,69 ^a	–	–	0,59	0,72 ^a	–	–	0,60
s6	–	0,53 ^a	–	–	0,40	0,57 ^a	–	–	0,34
s7	0,41	0,61	–	–	0,56	0,63 ^a	–	0,38	0,55
s8	–	0,70 ^a	–	–	0,63	0,71 ^a	0,33	–	0,65
s9	–	0,63 ^a	–	–	0,49	0,67 ^a	–	–	0,50
s10	–	0,63 ^a	–	–	0,53	0,64 ^a	–	–	0,50
s11	0,34	0,41 ^a	0,35	–	0,43	0,45 ^a	0,30	0,37	0,43
s12	–	0,65 ^a	–	–	0,54	0,65 ^a	–	–	0,51
s13	0,46	0,60	–	–	0,57	0,59	–	0,42	0,53
s14	0,55	0,47	–	–	0,53	0,52	–	0,52	0,54
s15	0,57	0,49	–	–	0,57	0,53	–	0,55	0,60
i1	0,34	–	0,66 ^a	–	0,57	–	0,58	0,40	0,50
i2	0,32	–	0,56 ^a	–	0,47	–	0,47 ^a	0,32	0,36
i3	–	–	0,64 ^a	0,38	0,60	–	0,73 ^a	–	0,58
i4	–	–	0,76 ^a	–	0,67	–	0,82 ^a	–	0,67
i5	–	–	0,78 ^a	–	0,69	–	0,85 ^a	–	0,73
i6	–	–	0,47 ^a	–	0,40	–	0,48 ^a	0,31	0,41
i7	–	–	0,39 ^a	–	0,26	–	0,42 ^a	–	0,26
i8	–	–	0,79 ^a	–	0,65	–	0,74 ^a	–	0,60
i9	–	–	0,82 ^a	–	0,69	–	0,81 ^a	–	0,65
i10	–	–	0,55 ^a	–	0,51	–	0,47	0,47	0,45
i11	0,66	–	–	–	0,48	–	–	0,67	0,49
i12	–	0,37	–	0,37	0,35	0,40	0,39	–	0,32
i13	–	–	0,39	0,50	0,41	–	0,54 ^a	–	0,32
i14	0,50	–	–	–	0,36	–	–	0,56	0,39
i15	0,41	–	0,31	–	0,33	–	0,31	0,48	0,36
k1	0,64 ^a	–	–	–	0,51	–	–	0,65 ^a	0,53
k2	0,64 ^a	–	–	–	0,52	–	–	0,65 ^a	0,51
k3	0,71 ^a	–	–	–	0,63	–	–	0,76 ^a	0,63
k4	0,57 ^a	–	–	0,30	0,49	–	–	0,63 ^a	0,47
k5	0,44	–	–	0,47	0,43	–	–	0,50 ^a	0,32
k6	–	–	–	0,57	0,40	–	–	0,35 ^a	0,22
k7	–	0,57	–	–	0,38	0,59	–	–	0,37
k8	–	–	–	0,69	0,50	–	–	–	0,13
k9	0,33	0,49	–	–	–	0,50	–	–	0,34
a1	–	–	–	0,65 ^a	0,52	–	–	–	–
a2	–	–	–	0,46 ^a	0,36	–	–	–	–
a3	–	–	–	0,52 ^a	0,36	–	–	–	–
a4	–	–	–	0,74 ^a	0,56	–	–	–	–
a5	–	–	–	0,63 ^a	0,40	–	–	–	–
a6	–	–	–	0,48 ^a	0,32	–	–	–	–
a7	0,36	–	–	0,40 ^a	0,37	–	–	–	–
a8	0,60	0,32	–	–	0,48	–	–	–	–
a9	0,66	–	–	–	0,49	–	–	–	–

Tab. 1 (Fortsetzung)

Item	4 Faktoren					3 Faktoren				
	a ₁	a ₂	a ₃	a ₄	<i>h</i> ²	a ₁	a ₂	a ₃	<i>h</i> ²	
a10	0,61	–	–	–	0,40	–	–	–	–	
a11	0,54	–	–	–	0,41	–	–	–	–	
a12	0,60	–	–	–	0,45	–	–	–	–	
a13	0,54	–	–	0,40	0,56	–	–	–	–	
a14	0,43	–	0,31	0,36	0,41	–	–	–	–	
a15	0,70	–	–	–	0,53	–	–	–	–	

Ladungen <0,30 ausgeblendet; Hauptkomponentenanalyse, rotierte Matrix (Varimax). *h*²: Kommunalität. a₁, a₂, a₃, a₄: Ladung der Items auf dem 1., 2., 3. oder 4. Faktor

s sozial, i individuell, k kriterial, a absolut

^aItem in weitere Auswertung (Reliabilitäten, Summenscores, Multiple Regression) einbezogen. Items mit unerwarteten Ladungen oder Querladungen ≥0,4 wurden verworfen

2003). Die Kommunalitäten der Einzelitems (vgl. Tab. 1) liegen überwiegend im mittleren bis vereinzelt hohen Bereich, was bei der vorliegenden Stichprobengröße einen Hinweis auf die Stabilität der Faktorenlösungen liefert. Einige niedrige Kommunalitäten kommen allerdings vor. Die Faktoren des vierfaktoriellen Ansatzes erklären zusammen 48,77 %, die Faktoren des dreifaktoriellen Ansatzes (unter Ausschluss der nicht-bezugsnormbeeinflussten Items) 47,17 % der Gesamtvarianz der Items.

In beiden durchgeführten EFA (vgl. Tab. 1) weist der überwiegende Anteil (s1 bis s13) der Items, die inhaltlich das von der sozialen Bezugsnorm belastete experimentbezogene FSK erfassen, die jeweils höchste Faktorladung auf einem gemeinsamen Faktor auf. Die Items s14 und 15 werden wegen unerwarteter Faktorladungen und die Items s1, s7, und s13 (4-Faktoren-EFA) bzw. s1 und s13 (3-Faktoren-EFA) wegen Querladungen ≥0,4 verworfen. Die verbleibenden zehn (4-Faktoren-EFA) bzw. elf Items (3-Faktoren-EFA) werden zu der Skala *sozial* zusammengefasst. Auch die Items i1 bis i10 (4-Faktoren-EFA) bzw. i1 bis i9 und i13 (3-Faktoren-EFA), welche das von der individuellen Bezugsnorm beeinflusste experimentbezogene FSK abbilden sollen, laden mit ihren höchsten Faktorladungen auf einen gemeinsamen Faktor. Die Items i11 bis i15 (4-Faktoren-EFA) bzw. i10 bis i12, i14 und i15 (3-Faktoren-EFA) werden wegen unklarer Faktorladungen, Item i1 (3-Faktoren-EFA) wird wegen einer Querladungen ≥0,4 verworfen. Die verbleibenden Items werden zu der Skala *individuell* zusammengefasst. Während der überwiegende Teil der kriterial belasteten Items (k1 bis k6) im dreifaktoriellen Ansatz ebenfalls deutlich auf einen gemeinsamen Faktor lädt, sind die Faktorladungen der Hälfte der kriterial belasteten Items im vierfaktoriellen Ansatz – unter Einbezug der nicht von Bezugsnormen belasteten Items (a1 bis a15) – nicht so konsistent. Zwar laden auch in der 4-Faktoren-EFA die Items k1 bis k4 klar auf einen gemeinsamen Faktor. Auf denselben Faktor laden aber auch weitere nicht bezugsnormbelastete Items (a8 bis a15). Ebenso laden einige kriterial belastete

Items (k5, k6, k8) gemeinsam mit den nicht-bezugsnormbelasteten Items a1 bis a7 auf den vierten Faktor. Da die Items k1 bis k4 im Wortlaut deutlich expliziter die kriteriale Bezugsnorm abbilden als die Items k5 bis k8 (z.B. k3: „Entstandene Messdaten des Experimentes kann ich so auswerten, dass sie den Anforderungen des Unterrichts gerecht werden“) und da die Items k1 bis k4 auf einen gemeinsamen Faktor laden, werden in die Konstruktion der Skala kriterial im vierfaktoriellen Ansatz lediglich diese Items einbezogen. Die bezugsnorm-unbelasteten Items, die ebenfalls mit k1 bis k4 auf einen gemeinsamen Faktor laden, werden verworfen, da sie inhaltlich keine kriteriale Bezugsnorm abbilden. Als Konsequenz werden im vierfaktoriellen Ansatz die Items a1 bis a7 in die weitere Konstruktion der Skala *absolut* einbezogen. Um später auch den Einfluss der Skalen des dreifaktoriellen Ansatzes auf das absolute experimentbezogene FSK untersuchen zu können wurden hier – der inhaltlichen Argumentation von Schöne et al. (2002) folgend – alle Items ohne Bezugsnormbelastung (a1 bis a15) in die Skalenkonstruktion einbezogen.

Trennschärfe und Reliabilität Aufgrund zufriedenstellender Trennschärfen (*h*² ≥ 0,30) und Reliabilitäten (Cronbachs $\alpha \geq 0,77$) mussten keine weiteren Items aus den Skalen ausgeschlossen werden (vgl. Tab. 2). Sowohl die aus dem vier- als auch dreifaktoriellen Ansatz hervorgehenden Skalen wurden hinsichtlich ihrer Eignung zur Durchführung multipler Regressionsanalysen getestet.

Multiple Regressionsanalysen In beiden Ansätzen deuten Streudiagramme (bezugsnormbelasteter Prädiktor vs. absolutes experimentbezogenes FSK) auf die Linearität der modellierten Zusammenhänge hin. Residuenbasierte Hinweise auf Ausreißer (Std. Residuen_{4 Faktor} < |2,44|; Std. Residuen_{3 Faktor} < |2,77|, vgl. Schlepper 2014; Hebelwerte_{4 Faktor} < 0,12, Hebelwerte_{3 Faktor} < 0,12, vgl. Huber 1981; Cook-Distanz_{4 Faktor} < 0,20, Cook-Distanz_{3 Faktor} ≤ 0,29, vgl. Kohn und Öztürk 2013), Heteroskedastizität (keine auf-

Tab. 2 Item- und Skalenkennwerte

Skala	α	Item	4 Faktoren			3 Faktoren		
			(Antwortskala: 0–4)			(Antwortskala: 0–4)		
			<i>M</i>	<i>SD</i>	<i>R</i> _{i,t-i}	<i>M</i>	<i>SD</i>	<i>R</i> _{i,t-i}
Sozial (s)	4 Faktoren $\alpha = 0,89$	s2	2,36	0,97	0,67	2,36	0,97	0,68
		s3	2,36	0,98	0,59	2,36	0,98	0,59
		s4	2,25	1,00	0,70	2,25	1,00	0,71
	3 Faktoren $\alpha = 0,90$	s5	2,23	0,93	0,71	2,23	0,93	0,72
		s6	2,46	0,84	0,46	2,46	0,84	0,47
		s7	–	–	–	2,08	0,90	0,63
		s8	2,41	0,97	0,62	2,41	0,97	0,61
		s9	2,59	1,02	0,65	2,59	1,02	0,65
		s10	2,32	1,02	0,65	2,32	1,02	0,66
		s11	2,46	1,00	0,57	2,46	1,00	0,58
		s12	2,11	0,99	0,66	2,11	0,99	0,66
		<i>Skala</i>	2,35	0,69	–	2,33	0,68	–
Individuell (i)	4 Faktoren $\alpha = 0,88$	i1	2,54	1,03	0,66	–	–	–
		i2	2,56	1,06	0,51	2,56	1,06	0,46
		i3	3,02	0,91	0,63	3,02	0,91	0,67
	3 Faktoren $\alpha = 0,86$	i4	2,86	0,99	0,71	2,86	0,99	0,71
		i5	2,92	1,06	0,71	2,92	1,06	0,71
		i6	2,74	1,03	0,50	2,74	1,03	0,47
		i7	2,89	0,85	0,41	2,89	0,85	0,42
		i8	2,92	1,00	0,68	2,92	1,00	0,67
		i9	3,02	1,00	0,71	3,02	1,00	0,72
		i10	2,32	0,99	0,57	–	–	–
		i13	–	–	–	3,09	1,01	0,47
		<i>Skala</i>	2,78	0,69	–	2,89	0,68	–
Kriterial (k)	4 Faktoren $\alpha = 0,77$	k1	2,08	0,97	0,46	2,08	0,97	0,47
		k2	2,64	0,88	0,54	2,64	0,88	0,57
		k3	2,57	0,88	0,72	2,57	0,88	0,71
	3 Faktoren $\alpha = 0,79$	k4	2,64	1,03	0,57	2,64	1,03	0,58
		k5	–	–	–	2,75	0,91	0,52
		k6	–	–	–	3,11	0,81	0,41
		<i>Skala</i>	2,48	0,72	–	2,63	0,64	–
Absolut (a)	4 Faktoren $\alpha = 0,77$	a1	2,92	0,91	0,63	2,92	0,91	0,55
		a2	2,90	0,86	0,46	2,90	0,86	0,49
		a3	3,22	0,67	0,41	3,22	0,67	0,43
	3 Faktoren $\alpha = 0,86$	a4	3,08	0,89	0,57	3,08	0,89	0,47
		a5	3,45	0,71	0,40	3,45	0,71	0,30
		a6	3,18	0,84	0,48	3,18	0,84	0,46
		a7	3,02	0,88	0,49	3,02	0,88	0,54
		a8	–	–	–	2,28	0,91	0,52
		a9	–	–	–	2,12	1,01	0,54
		a10	–	–	–	2,61	0,89	0,47
		a11	–	–	–	2,39	0,92	0,55
		a12	–	–	–	2,49	0,87	0,57
		a13	–	–	–	2,91	0,93	0,67
		a14	–	–	–	2,94	1,00	0,45
		a15	–	–	–	2,21	0,84	0,54
		<i>Skala</i>	3,11	0,54	–	2,78	0,52	–

N = 166, *r*_{i,t-i}: Trennschärfe

fällig trichterförmige Verteilung im Streudiagrammen der standardisierten Residuen vs. vorhergesagte Werte, vgl. Tabachnick und Fidell 2014) oder Abweichungen der Residuen von der Normalverteilung ($D_{4 \text{ Faktor}}(166)=0,05$, $p=0,200$, $D_{3 \text{ Faktor}}(166)=0,07$, $p=0,084$, vgl. Janssen und Laatz 2017) liegen in keinem der beiden Ansätze vor. Die Durbin-Watson-Statistik weist nicht auf extreme Autokorrelationen hin ($\text{Durbin-Watson}_{4 \text{ Faktor}}=1,82$, $\text{Durbin-Watson}_{3 \text{ Faktor}}=1,65$, vgl. Field 2013). Moderate Korrelationen der Prädiktoren (vgl. Tab. 3) sowie die VIF-Werte < 2 ($\text{VIF}_{4 \text{ Faktor}} \leq 1,52$, $\text{VIF}_{3 \text{ Faktor}} \leq 1,55$, vgl. Schneider 2007) liefern keinen Hinweis auf das Vorliegen von Multikollinearität. Die Durchführung multipler Regressionen erscheint daher zulässig.

Das postulierte und mittels ANOVA überprüfte Regressionsmodell (Prädiktoren: kriteriales, soziales und individuelles experimentbezogenes FSK; abhängige Variable: absolutes experimentbezogenes FSK) wird sowohl von den Daten aus dem vier- als auch dreifaktoriellen Ansatz gestützt (vgl. Tab. 4). Die gewählten Prädiktoren klären einen hohen Anteil (vgl. Cohen 1988) der Varianz des absoluten experimentbezogenen FSK auf. 40 % (Skalen aus dem 4-Faktoren-Ansatz) bzw. 75 % (Skalen aus dem 3-Faktoren-Ansatz) der Variation des absoluten experimentbezogenen FSK können durch den Einfluss der drei Bezugsnormbelasteten FSK erklärt werden. Während bei Einsatz der Skalen aus dem vierfaktoriellen Ansatz alle Prädiktoren einen mindestens signifikanten ($p \leq 0,05$) Einfluss auf das absolute experimentbezogene FSK haben, weisen bei Einsatz der Skalen aus dem dreifaktoriellen Ansatz lediglich der von der kriterialen sowie sozialen Bezugsnorm belastete Prädiktor einen signifikanten Einfluss ($p \leq 0,001$) auf die abhängige Variable auf. Der Einfluss der individuellen experimentbezogenen FSK wird nur tendenziell ($p < 0,10$) gestützt. Beim Vergleich der standardisierten Regressionskoeffizienten wird deutlich, dass sowohl im vier- als auch dreifaktoriellen Ansatz das kriteriale experimentbezogene FSK den deutlich höchsten Einfluss auf das absolute experimentbezogene FSK hat. Der Einfluss der beiden anderen Prädiktoren fällt deutlich geringer aus. Während der Einfluss des sozialen und individuellen experimentbezogenen FSK im vierfaktoriellen Ansatz in Relation zueinander ungefähr gleich hoch ausfallen, ist ein Einfluss des individuellen experimentbezogenen FSK bei Einsatz der Skalen aus dem dreifaktoriellen Ansatz nicht zu belegen (vgl. Tab. 4).

Diskussion

Ziel dieser Studie war es, die von Schüler*innen zur Beurteilung des experimentbezogenen Fähigkeitsselbstkonzepts herangezogenen Bezugsnormen zu untersuchen. Die als Konsequenz der Faktoren- und Reliabilitätsanalyse beschriebene Itemreduzierung führte zu einem Messinstrument, welches in vier trennscharfen, reliablen Skalen das absolute, sowie das von der sozialen, kriterialen und individuellen Bezugsnorm beeinflusste experimentbezogene Fähigkeitsselbstkonzept erfasst. Vor dem Hintergrund des notwendigen Ausschlusses relativ vieler Items der Skalen absolut und kriterial erscheint dennoch eine Überarbeitung der ausgeschlossenen Items für zukünftige Studien sinnvoll, um eine vollständigere inhaltliche Abdeckung der beiden Konstrukte und damit eine höhere inhaltliche Validität zu gewährleisten. Vielversprechend erscheint der Ansatz einer expliziteren Repräsentation der kriterialen Bezugsnorm in den ausgeschlossenen Items der Skala kriterial. Denn wie die Faktorladungen zeigen, trat das Problem der gemeinsamen Faktorladungen nicht bei den Items auf, in denen die Bezugsnorm bereits sehr explizit – im Sinne von „den Anforderungen des Unterrichts gerecht werden“ – abgebildet wurde.

Wie die Ergebnisse der beiden multiplen Regressionsanalysen übereinstimmend – unabhängig davon, ob die Skalen aus der drei- bzw. vierfaktoriellen Faktorenanalyse in die Auswertung einbezogen wurden – andeuten, ziehen die untersuchten Schüler*innen insbesondere kriteriale Vergleiche mit den Ansprüchen des Systems Schule heran, um zur Einschätzung ihrer absoluten Experimentierfähigkeit zu gelangen. Der Einfluss der sozialen und individuellen Bezugsnorm fällt hingegen wesentlich geringer aus. Dies überrascht zunächst, da im schulischen Umfeld häufig die sozialen Vergleiche als Faktor für die Genese akademischer Selbstkonzepte gelten (vgl. Helmke 1992; Wagner 1999). Auch der in verschiedenen Unterrichtskontexten zu beobachtende Fischteichereffekt deutet die Relevanz sozialer Vergleiche an (vgl. Marsh und Hau 2003). Der in der vorliegenden Studie festgestellte moderate Einfluss sozialer Vergleiche auf das absolute experimentbezogene FSK kann (und soll) aufgrund des Untersuchungsdesigns eine Relevanz des Fischteichereffektes für die Höhe des experimentbezogenen FSK aber nicht ausschließen. Zukünftige Studien wären unter Berücksichtigung der tatsächlichen indivi-

Tab. 3 Interkorrelationen der Skalen des experimentbezogenen FSK

	Kriterial		Sozial		Individuell	
	4 Faktoren	3 Faktoren	4 Faktoren	3 Faktoren	4 Faktoren	3 Faktoren
Sozial	0,50	0,52	–	–	–	–
Individuell	0,34	0,37	0,46	0,46	–	–
Absolut	0,57	0,84	0,51	0,63	0,42	0,43

Tab. 4 Multiple Regressionen zur Prädiktion des absoluten experimentbezogenen Fähigkeitsselbstkonzepts durch die bezugsnormbelasteten Fähigkeitsselbstkonzepte, β : (standardisierte) Regressionskoeffizienten, S.E.: Standardfehler; Einbeziehung der sich aus den Faktorenanalysen mit vier bzw. drei vorgegebenen Faktoren ergebenden Skalen (vgl. Tab. 1)

Ansatz	4 Faktoren			3 Faktoren		
	β	β_{stand}	S.E.	β	β_{stand}	S.E.
Konstante	1,58	–	–	0,74	–	0,10
Kriterial	0,29***	0,39***	0,05	0,56***	0,69***	0,04
Sozial	0,18**	0,23**	0,06	0,18***	0,23***	0,04
Individuell	0,14*	0,18*	0,05	0,06 ^{tend}	0,08 ^{tend}	0,03
R^2	0,41			0,76		
Korr. R^2	0,40			0,75		
F-Test	$F(3,162) = 37,81***$			$F(3,162) = 169,27***$		

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$, tend: $p < 0,1$

duellen Experimentierfähigkeit der Schüler*innen als Maß der Experimentierfähigkeit des Referenzfeldes zwecks Beurteilung der Relevanz des Fischteich effektes wünschenswert. In diesem Kontext wäre es auch von Interesse zu untersuchen, ob die realisierten sozialen Vergleiche primär auf Personen mit höherer Experimentierfähigkeit gerichtet sind, wie es die von Festinger (1954) postulierte Tendenz von Aufwärtsvergleichen im Falle von Fähigkeitsvergleichen beschreibt. Hier wäre zudem zu prüfen, ob nicht gerade ein experimentierstarkes Lernumfeld durch die in diesem sozialen Umfeld vorherrschenden normativ differenzierten Lern- und Entwicklungsbedingungen – entgegen der Beobachtungen des Fischteich effektes – größere individuelle Leistungsfortschritte erlaubt (vgl. Jerusalem 1997; Zurbirgen 2016). Eine solche individuelle Steigerung der Experimentierleistung ließe einen positiven Einfluss auf das experimentbezogene FSK erwarten, denn wie die vorliegende Studie zeigt, deutet sich ein – wenn auch moderater – positiver Einfluss der individuellen Bezugsnorm auf das experimentbezogene FSK an.

Eine Erklärung für die beobachtete Dominanz des Einflusses der kriterialen Bezugsnorm könnte in der Neutralität – der Personenunabhängigkeit – des kriterialen Bezugsrahmens (Lohbeck 2017) sowie im Gerechtigkeitsempfinden der Schüler*innen zu suchen sein: Schüler*innen empfinden die Bewertung durch eine Lehrkraft dann als besonders gerecht, wenn sie anhand einer kriterialen, nicht aber anhand einer sozialen Bezugsnorm der Lehrkraft erfolgt (Dalbert et al. 2007). Diese Empfindung der Schüler*innen deckt sich mit den Vorgaben der Bewertungsrichtlinien für Schüler*innenleistungen der Länder. So sieht §48 III SchulG NRW im Rahmen der Definition der Schulnoten eine Orientierung der Leistungsbewertung durch die Lehrkräfte an den Anforderungen, also an den kriterialen Ansprüchen des Systems Schule vor, welche sich u. a. in den schulischen Curricula manifestieren. Zudem soll allerdings die Leistungsbewertung auch den Stand des Lernprozesses der Schüler*innen abbilden (vgl. §48 I S.1 SchulG

NRW), also die individuelle Entwicklung berücksichtigen. Sollte sich die Präferenz der Schüler*innen für bestimmte Bezugsnormen tatsächlich mit den von Lehrkräften im Rahmen der Leistungsbewertung herangezogenen Bezugsnormen decken, wäre in der vorliegenden Studie somit ein höherer Einfluss der individuellen Bezugsnorm zu erwarten gewesen. Rheinberg et al. (1977) stellen zudem fest, dass die Bezugsnormorientierung von Schüler*innen einem altersabhängigen Wandel unterliegt und dass insbesondere Schüler*innen der gymnasialen Oberstufe – wie sie in der vorliegenden Studie untersucht wurden – sich stärker an individuellen als an sozialen Bezugsnormen orientieren, im Vergleich zu Schüler*innen der Sekundarstufe I (Rheinberg et al. 1977). Eine Erklärung für den moderaten Einfluss der individuellen Bezugsnorm auf das experimentbezogene FSK in der vorliegenden Studie könnte in der relativ geringen Experimentiererfahrung vieler Schüler*innen (vgl. Seidel et al. 2007) – insbesondere das selbstständige Planen und Auswerten von Experimenten wird im naturwissenschaftlichen Unterricht nur wenig verlangt (vgl. Tesch und Duit 2004) – sowie in der wissenschaftsfernen Vorstrukturiertheit des Experimentalunterrichts (vgl. Höttecke und Rieß 2015) zu suchen sein. Die fehlenden Erfolgserlebnisse beim Experimentieren und die mangelnden Gelegenheiten Experimentierfehler zu begehen geben Schüler*innen möglicherweise nicht hinreichend Anlass zur Reflexion der eigenen Lernfortschritte (vgl. Gruber et al. 2006). Entsprechend setzen sie ihre heutigen Leistungen im Planen, Durchführen und Auswerten von Experimenten nicht zu ihren früheren Experimentierleistungen in Bezug.

Der in der vorliegenden Studie beobachtete deutliche Einfluss der kriterialen Bezugsnorm auf das absolute experimentbezogene FSK hat für die Praxis des Experimentalunterrichts hohe Relevanz. Die Schüler*innen gleichen demnach primär ab, ob sie den kriterialen Anforderungen des Experimentalunterrichts gerecht werden, um zur Einschätzung ihrer Experimentierfähigkeit zu gelangen. Da nur wenige Schüler*innen die schulischen Curricula zur Normie-

rung schulischer Leistungsansprüche heranziehen dürften, betonen die vorliegenden Ergebnisse – in Übereinstimmung mit Hattie (2009) – die besondere Bedeutung der Person der Lehrkraft: Durch Auswahl geeigneter Experimentieraufgaben, Experimentieranleitungen und Experimentiermaterialien in Orientierung an curricularen Vorgaben und Zielen sowie durch individuelles Feedback zur Experimentierfähigkeit im praktischen naturwissenschaftlichen Experimentierunterricht vermitteln Lehrkräfte ihren Schüler*innen täglich die kriterialen experimentbezogenen Ansprüche des Systems Schule. Dass Feedback der Lehrkraft einen deutlichen Einfluss auf unterrichtliche Leistung der Schüler*innen hat (vgl. Hattie und Timperley 2007; Hattie 2009), welcher gegebenenfalls über FSK moderiert wird (Baadte und Schnotz 2014), ist anzunehmen. Fraglich ist aber, welche Form des Feedbacks besonders geeignet erscheint, um Schüler*innen im Experimentalunterricht zu einem positiven und gleichermaßen realistischen experimentbezogenen FSK zu begleiten. Insbesondere für attributionales Feedback, welches unterrichtliche Erfolge auf die persönlichen Begabungen von Schüler*innen zurückführt, ist ein positiver Einfluss auf unterrichtliche FSK belegt (Schunk 1982; Craven et al. 1991; Penn et al. 2001). Zwar überwiegt im Unterricht die Quantität des Feedbacks durch Mitschüler*innen gegenüber dem Feedback durch die Lehrkraft (vgl. Nuthall 2007), dennoch scheinen Schüler*innen Feedback einer Lehrkraft eher zu adaptieren (Miao und Badger 2006). Studien zum Einfluss des Feedbacks der Lehrkraft als kriterialer Vermittler der (Experimentier-)Ansprüche des Systems Schule auf das experimentbezogene FSK der Schüler*innen erscheinen vor diesem Hintergrund von Interesse.

Auch wenn in beiden Regressionsanalysen die besondere Relevanz der kriterialen Bezugsnorm für die Höhe des absoluten experimentbezogenen FSK von Schüler*innen deutlich wird, fällt doch ein Unterschied in der Varianzaufklärung und in der absoluten Einflusshöhe der individuellen sowie der kriterialen Bezugsnorm zwischen beiden Ansätzen auf. Die sehr hohe Varianzaufklärung unter Einbezug der Skalen aus der 3-Faktoren-EFA erklärt sich aus dem besonders hohen Einfluss der Skala kriterial in diesem Ansatz. Eine wesentliche Ursache dafür ist in dem vollständigen Einbezug aller Items der Skala absolut in die – auf den Skalen der 3-Faktoren-EFA basierende – Regressionsanalyse zu sehen. Die Items a8–a15 der Skala absolut laden nämlich (wie deren Faktorladungen in der 4-Faktoren-EFA zeigen) deutlich mit den Items der Skala kriterial auf einen gemeinsamen Faktor, was unweigerlich bei deren Einbezug in die Regressionsanalyse (wie in dem 3-Faktoren-Ansatz gegeben) zu einer deutlichen Erhöhung des prädiktiven Charakters der Skala kriterial und zu einer Gefahr der Überschätzung ihres Einflusses führt. Dieser Gefahr begegnet der berichtete 4-Faktoren-Ansatz durch Ausschluss der problematischen Items a8–a15, so dass die absolute Höhe

des Einflusses der kriterialen Bezugsnorm hier ein realistischeres Bild liefert. Entsprechend erklärt sich auch der geringere Einfluss der individuellen Bezugsnorm im 3-Faktoren-Ansatz aus dem eventuell überschätzten Einfluss der kriterialen Bezugsnorm in diesem Ansatz, denn dieser lässt – da die multiple Regressionsanalyse relative Bedeutungen der Prädiktoren aufzeigt (vgl. Budescu 1993) – die relative Relevanz der individuellen Bezugsnorm hier besonders gering erscheinen. Entsprechend sinnvoll erscheint – analog zum Vorgehen bei Sparfeldt et al. (2003) – der Einbezug der Items der Skala absolut in die EFA, wie dieser in dem berichteten 4-Faktoren-Ansatz vorgenommen wurde. Auch das hierauf basierende multiple Regressionsmodell weist mit 40 % noch eine hohe Varianzaufklärung auf (vgl. Cohen 1988). Diese deutet allerdings auch an, dass noch weitere Einflussfaktoren für das absolute experimentbezogene FSK relevant sind.

Diese gilt es in zukünftigen Studien zu klären. So sollte insbesondere die Experimentiererfahrung erfasst werden, da Schüler*innen vergangene Ereignisse heranziehen, um ihre zukünftigen Erfolgserwartungen zu generieren (vgl. Faber 2012). Es erscheint allerdings wenig sinnvoll, ausschließlich die bisherige Experimentierhäufigkeit zu erfassen. Vielmehr wirken auch die subjektiven Interpretationen vorausgegangener Erfahrungen, vermittelt über die daraus resultierenden affektiven Reaktionen und Erinnerungen auf das Konzept eigener Fähigkeiten (vgl. Erwartungs-Wert-Modell, Eccles et al. 1983). Somit erscheint es relevant, die aus den positiven oder negativen Erfahrungen in Experimentiersituationen resultierenden Ursachenzuschreibungen für Erfolge bzw. Misserfolge beim Experimentieren als potentielle Einflussgrößen auf das absolute experimentbezogene FSK zu berücksichtigen. So ist in Orientierung am Schema der Kausalattributionen nach Weiner (1988) davon auszugehen, dass Schüler*innen in unterschiedlichem Ausmaß internale (Begabung und Anstrengung) sowie externale (Aufgabenschwierigkeit und Zufall) Ursachen für die Erklärung von unterrichtlichen Experimentiererfolgen bzw. -misserfolgen heranziehen. Bereits in anderen unterrichtlichen Kontexten wurden Zusammenhänge von Kausalattributionen und FSK belegt (Marsh 1984). Weiterhin wäre für zukünftige Studien eine differenziertere Erfassung der individuellen Bezugsnorm in Erwägung zu ziehen. Zwar ist die individuelle Bezugsnorm – so auch im vorliegenden Messinstrument – klassischerweise definiert als der temporale intrapersonale Vergleich der aktuellen eigenen akademischen Fähigkeiten zu früheren Fähigkeiten (vgl. Rheinberg 2001), Untersuchungen auf Basis des internal/external frame of reference models belegen darüber hinaus aber die Relevanz individueller interfachlicher Vergleiche für FSK (vgl. Skaalvik und Rankin 1992; Möller et al. 2009). Neben den externen (= sozialen) Vergleichen werden hier internale Vergleiche der wahrgenommenen Fähigkeiten in einem Gebiet

(z. B. Mathematik) mit den wahrgenommenen Fähigkeiten in einem anderen Gebiet (z. B. Sprachen) als relevante Größen für die Ausbildung fachspezifischer FSK gesehen (vgl. Skaalvik und Rankin 1992). Entsprechend sinnvoll wäre es daher, in kommenden Studien die Erfassung der individuellen Bezugsnorm auszuweiten, indem – neben temporalen Vergleichen mit der früheren Experimentierfähigkeit – zusätzlich Vergleiche zu anderen Fähigkeitsgebieten (z. B. Vergleiche zu sprachlichen, allgemein theoretischen bzw. allgemein praktischen Fähigkeiten) berücksichtigt werden. Eine Erweiterung des vorliegenden Modells um die beschriebenen Einflussgrößen könnte einen weiteren Beitrag zur Aufklärung der Varianz des experimentbezogenen Fähigkeitsselbstkonzepts leisten.

Funding Open Access funding provided by Projekt DEAL.

Interessenkonflikt R. Atzert, R. John, A. Preisfeld und K. Damerau erklären, dass keine Interessenkonflikte im Zusammenhang mit diesem Artikel bestehen.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Baadte, C., & Schnotz, W. (2014). Feedback effects on performance, motivation and mood: are they moderated by the learner's self-concept? *Scandinavian Journal of Educational Research*, 58(5), 570–591.
- Bartlett, M. S. (1951). The Effect of Standardization on a Chi Square Approximation in Factor Analysis. *Biometrika*, 38, 337–344.
- Barzel, B., Reinhoffer, B., & Schrenk, M. (2012). Das Experiment im Unterricht. In W. Rieß, M. Wirtz, B. Barzel & A. Schulz (Hrsg.), *Experimentieren im mathematisch-naturwissenschaftlichen Unterricht – Schüler lernen wissenschaftlich denken und arbeiten*. Münster, New York, München, Berlin: Waxmann.
- Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research, Evaluation*, 18(6), 1–13.
- Brunstein, J. C., & Heckhausen, H. (2018). Leistungsmotivation. In J. Heckhausen & H. Heckhausen (Hrsg.), *Motivation und Handeln* (S. 163–221). Berlin, Heidelberg: Springer.
- Budescu, D. V. (1993). Dominance analysis: a new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, 114(3), 542–551.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.). München: Pearson.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, N.J.: Erlbaum.
- Craven, R. G., Marsh, H. W., & Debus, R. L. (1991). Effects of internally focused feedback and attributional feedback on enhancement of academic self-concept. *Journal of Educational Psychology*, 83, 17–27.
- Dalbert, C., Schneidewind, U., & Saalbach, A. (2007). Justice judgments concerning grading in school. *Contemporary Educational Psychology*, 32, 420–433.
- Denissen, J. J., Zaretz, N. R., & Eccles, J. S. (2007). Do it, i'm able, and i know i am: Longitudinal couplings between domain-specific achievement, self-concept, and interest. *Child Development*, 78(2), 430–447.
- Dickhäuser, O. (2006). Fähigkeitsselbstkonzepte – Entstehung, Auswirkung, Förderung. *Zeitschrift für Pädagogische Psychologie*, 20, 5–8.
- Dickhäuser, O., & Rheinberg, F. (2003). Bezugsnormorientierung: Erfassung, Probleme, Perspektiven. In J. Stiensmeier-Pelster & F. Rheinberg (Hrsg.), *Diagnostik von Motivation und Selbstkonzept*. Göttingen: Hogrefe.
- Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., & Meece, J. L. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Hrsg.), *Achievement and achievement motives: Psychological and sociological approaches* (S. 75–146). San Francisco: W. H. Freeman.
- Ertl, B., Luttenberger, S., & Paechter, M. (2014). Stereotype als Einflussfaktoren auf die Motivation und die Einschätzung der eigenen Fähigkeiten bei Studentinnen in MINT-Fächern. *Gruppendynamik und Organisationsberatung*, 45(4), 419–440.
- Faber, G. (2012). *Selbstkonzept, Kausalattribution und Leistungsangst im Rechtschreiben. Schulfachspezifische Analysen zu den Selbst-einschätzungen von Grundschulkindern*. Saarbrücken: AV.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117–140.
- Field, A. (2013). *Discovering statistics using IBM SPSS Statistics*. London: SAGE.
- Fisseni, H. J. (1997). *Lehrbuch der psychologischen Diagnostik*. Göttingen: Hogrefe.
- Fromm, S. (2012). Faktoren- und Reliabilitätsanalyse. In S. Fromm (Hrsg.), *Datenanalyse mit SPSS für Fortgeschrittene 2: Multivariate Verfahren für Querschnittsdaten* (S. 53–82). Wiesbaden: Springer VS.
- Götz, T., & Preckel, F. (2006). Der „Big-fish-little-pond-Effekt“ („Fischteicheffekt“): Eine Untersuchung an der Sir-Karl-Popper-Schule und am Wiedner Gymnasium in Wien. *Özbf news & science*, 14, 24–26.
- Gruber, H., Harteis, C., & Rehr, M. (2006). Professional Learning: Erfahrung als Grundlage von Handlungskompetenz. *Bildung und Erziehung*, 59(2), 193–204.
- Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research*, 52(1), 123–142.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Hattie, J. (2009). *Visible learning – A synthesis of over 800 meta-analyses relating to achievement*. London, New York: Routledge.
- Heckhausen, H. (1974). *Leistung und Chancengleichheit*. Göttingen: Hogrefe.
- Heinicke, S., & Peters, S. (2014). Was ist Experimentieren? – Populäre Sichtweisen unter der Lupe. *Unterricht Physik*, 144, 10–13.
- Helmke, A. (1992). *Selbstvertrauen und schulische Leistungen*. Göttingen: Hogrefe.
- Höttecke, D., & Rieß, F. (2015). Naturwissenschaftliches Experimentieren im Lichte der jüngeren Wissenschaftsforschung – Auf der Suche nach einem authentischen Experimentbegriff der Fachdi-

- daktik. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 127–139.
- Huber, P.J. (1981). *Robust Statistics*. New York: Wiley.
- Jäger, R.S. (2008). Leistungsbeurteilung. In W. Schneider & M. Haselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (S. 324–336). Göttingen: Hogrefe.
- Janssen, J., & Laatz, W. (2017). *Statistische Datenanalyse mit SPSS – Eine anwendungsorientierte Einführung in das Basissystem und das Modul Exakte Tests* (9. Aufl.). Berlin, Heidelberg: Springer Gabler.
- Jerusalem, M. (1997). Schulklasseneffekte. In F.E. Weinert (Hrsg.), *Enzyklopädie der Psychologie: Psychologie des Unterrichts und der Schule* (Bd. 3, S. 253–278). Göttingen: Hogrefe.
- Kaiser, H.F., & Rice, J. (1974). Little Jiffy – Mark IV. *Educational and Psychological Measurement*, 34, 111–117.
- Kohn, W., & Öztürk, R. (2013). *Statistik für Ökonomen – Datenanalyse mit R und SPSS* (2. Aufl.). Berlin, Heidelberg: Springer Gabler.
- Köller, O., Daniels, Z., Schnabel, K.U., & Baumert, J. (2000). Kurswahlen von Mädchen und Jungen im Fach Mathematik: Zur Rolle von fachspezifischem Selbstkonzept und Interesse. *Zeitschrift für Pädagogische Psychologie*, 14(1), 26–37.
- Köller, O., Trautwein, U., Lüdtke, O., & Baumer, J. (2006). Zum Zusammenspiel von schulischer Leistung, Selbstkonzept und Interesse in der gymnasialen Oberstufe. *Zeitschrift für Pädagogische Psychologie*, 20(1/2), 27–39.
- Little, R.J.A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83, 1198–1202.
- Lohbeck, A. (2017). Die individuell präferierte Bezugsnormorientierung und das Selbstkonzept von Grundschulkindern im Fach Mathematik. *Zeitschrift für Pädagogische Psychologie*, 31, 41–55.
- Mabe, P.A., & West, S.G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of applied Psychology*, 67(3), 280–296.
- Marsh, H.W. (1984). Relations among dimensions of self-attribution, dimensions of self-concept, and academic achievements. *Journal of Educational Psychology*, 76(6), 1291–1308.
- Marsh, H.W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, 79(3), 280–295.
- Marsh, H.W., & Hau, K.-T. (2003). Big-Fish—Little-Pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist*, 58(5), 364–376.
- Marsh, H.W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational psychologist*, 20(3), 107–123.
- Marsh, H.W., Shavelson, R.J., & Byrne, B.M. (1988). A multifaceted academic selfconcept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology*, 80(3), 66–80.
- Meyer, W.U. (1973). *Leistungsmotiv und Ursachenerklärung von Erfolg und Mißerfolg*. Stuttgart: Klett.
- Meyer, W.-U. (1984). *Das Konzept von der eigenen Begabung*. Bern: Huber.
- Miao, Y., & Badger, R. (2006). A comparative study of peer and teacher feedback in a chinese EFL writing class. *Journal of second language writing*, 15(3), 179–200.
- Mischo, C., & Rheinberg, F. (1995). Erziehungsziele von Lehrern und individuelle Bezugsnormen der Leistungsbewertung. *Zeitschrift für Pädagogische Psychologie*, 9(3/4), 139–151.
- Möller, J., Pohlmann, B., Köller, O., & Marsh, H.W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research*, 79(3), 1129–1167.
- Noormann, P. (2017). *Mehrstufige Eigenmarken – Eine empirische Analyse von Zielen, Erfolgsdeterminanten und Grenzen*. Wiesbaden: Springer Gabler.
- Nuthall, G.A. (2007). *The hidden lives of learners*. Wellington: New Zealand Council for Educational Research.
- Penn, C. S., Burnett, P. C., & Patton, W. (2001). The impact of attributional feedback on the self-concept of children aged four to six years in preschool. *Journal of Psychologists and Counsellors in Schools*, 11, 21–34.
- Priess-Groben, H.A., & Hyde, J.S. (2017). Implicit theories, expectancies, and values predict mathematics motivation and behavior across high school and college. *Journal of youth and adolescence*, 44(6), 1318–1332.
- Ratschinski, G. (2009). *Selbstkonzept und Berufswahl*. Münster: Waxmann.
- Rheinberg, F. (2001). Bezugsnormen und schulische Leistungsbeurteilung. In F.E. Weinert (Hrsg.), *Leistungsmessung in Schulen* (S. 59–71). Weinheim: Beltz.
- Rheinberg, F., Lüthmann, J.V., & Wagner, H. (1977). Bezugsnorm-Orientierung von Schülern der 5. bis 13. Klasse bei der Leistungsbeurteilung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 2, 90–93.
- Ruploh, B., Martzy, F., Bischoff, A., Matschulat, N., & Zimmer, R. (2014). Veränderungen im Selbstkonzept nach psychomotorischer Förderung – Eine Pilotstudie im Mixed Methods. *Design. motorik*, 36(4), 180–189.
- Schiepe-Tiska, A., & Schmidtner, S. (2013). Mathematikbezogene emotionale und motivationale Orientierungen, Einstellungen und Verhaltensweisen von Jugendlichen in PISA 2012. In M. Pränzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012 – Fortschritte und Herausforderungen in Deutschland* (S. 99–121). Münster: Waxmann.
- Schiepe-Tiska, A., Simm, I., & Schmidtner, S. (2016). Motivationale Orientierungen, Selbstbilder und Berufserwartungen in den Naturwissenschaften in PISA 2015. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (S. 99–132). Münster, New York: Waxmann.
- Schlepper, F. (2014). *Vermarktung von Sportveranstaltungen*. Wiesbaden: Springer.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350–353.
- Schneider, H. (2007). Nachweis und Behandlung von Multikollinearität. In S. Albers, D. Klapper, U. Konradt, A. Walter & J. Wolf (Hrsg.), *Methodik der empirischen Forschung* (2. Aufl. S. 183–198). Wiesbaden: Gabler.
- Schöne, C., Dickhäuser, O., Spinath, B., & Stiensmeier-Pelster, J. (2002). *Skalen zur Erfassung des schulischen Selbstkonzepts (SESSKO)*. Göttingen: Hogrefe.
- Schöne, C., Dickhäuser, O., Spinath, B., & Stiensmeier-Pelster, J. (2003). Das Fähigkeitsselbstkonzept und seine Erfassung. In J. Stiensmeier-Pelster & F. Rheinberg (Hrsg.), *Diagnostik von Motivation und Selbstkonzept* (S. 3–14). Göttingen: Hogrefe.
- Schöne, C., Dickhäuser, O., Spinath, B., & Stiensmeier-Pelster, J. (2004). Zielorientierung und Bezugsnormorientierung: Zum Zusammenhang zweier Konzepte. *Zeitschrift für Pädagogische Psychologie*, 18, 93–99. <https://doi.org/10.1024/1010-0652.18.2.93>.
- Schöne, C., Dickhäuser, O., Spinath, B., & Stiensmeier-Pelster, J. (2012). *SESSKO – Skalen zur Erfassung des schulischen Selbstkonzepts* (Bd. 2, überarbeitete und neu normierte Auflage). Göttingen: Hogrefe.
- Schreiber, N., Theyßen, H., & Schecker, H. (2009). Experimentelle Kompetenz messen?! *Physik und Didaktik in Schule und Hochschule*, 3(8), 92–101.
- Schreiber N., Theyßen H., Dickmann. (2016). Wie genau beurteilen Schülerinnen und Schüler ihre eigenen experimentellen Fähigkeiten? – Ein Ansatz zur praktikablen Diagnostik experimenteller Fähigkeiten im Unterrichtsalltag. *Physik und Didaktik in Schule und Hochschule*, 1(15), 49–63.
- Schunk, D.H. (1982). Effects of effort attributional feedback on children's perceived self-efficacy and achievement. *Journal of Educational Psychology*, 74, 548–556.

- Seidel, T., Prenzel, M., Wittwer, J. & Schwindt, K. (2007). Unterricht in den Naturwissenschaften. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 147–180). Münster: Waxmann.
- Shavelson, R. J., J., H. J., & Stanton, J. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, 46, 407–441.
- Skaalvik, E. M., & Rankin, R. J. (1992). Math and verbal achievement and self-concepts: Testing the internal/external frame of reference model. *Journal of Early Adolescence*, 12(3), 267–279.
- Sparfeldt, J. R., Schilling, S. R., Rost, D. H., & Müller, C. (2003). Bezugsnormorientierte Selbstkonzepte? Zur Eignung der SESS-KO. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 24(4), 325–335.
- Spinath, B. (2011). Lernmotivation. In H. Reinders, H. Ditton, C. Gräsel & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung* (S. 55–68). Wiesbaden: Springer VS.
- Stiensmeier-Pelster, J., & Schöne, C. (2008). Fähigkeitsselbstkonzept. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (S. 62–73). Göttingen: Hogrefe.
- Sülz, C. (2014). *Bezugsnormen in der Leistungsmotivation* (Diss.). Greifswald: Ernst-Moritz-Arndt-Universität.
- Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics*. Harlow: Pearson.
- Tesch, M., & Duit, R. (2004). Experimentieren im Physikunterricht – Ergebnisse einer Videostudie. *Zeitschrift für Didaktik der Naturwissenschaften*, 10(10), 51–69.
- Wagner, J. W. L. (1999). *Soziale Vergleiche und Selbsteinschätzungen*. Münster: Waxmann.
- Weiner, B. (1988). *Motivationspsychologie*. Weinheim: Beltz Psychologie Verlags Union.
- Zurbriggen, C. (2016). Die schulische Bezugsgruppe als soziales Lernumfeld. In C. Zurbriggen (Hrsg.), *Schulklasseneffekte* (S. 49–90). Wiesbaden: Springer VS.