

## **Kurzskala zur Messung gemäßigt konstruktivistischer Prozessmerkmale (Kurz-PgK) im Biologieunterricht**

Melanie Basten, Samuel Greiff, Sabine Marsch, Annika Meyer, Detlef Urhahne & Matthias Wilde

melanie.basten@uni-bielefeld.de – samuel.greiff@uni.lu –  
detlef.urhahne@uni-passau.de – matthias.wilde@uni-bielefeld.de

Universität Bielefeld - Biologiedidaktik,  
Universitätsstr. 25, 33615 Bielefeld

–  
Institute of Cognitive Science and Assessment - ECCS unit  
6, rue Richard Coudenhove Kalergi, 1359 Luxembourg-Kirchberg

–  
Universität Passau - Pädagogische Psychologie  
Innstraße 40, 94032 Passau

---

### **Zusammenfassung**

*Im Biologieunterricht gewinnen konstruktivistische Lernumgebungen zunehmend an Bedeutung. Jedoch liegt bislang keine zeitökonomische Methode zur Erfassung konstruktivistischer Prozessmerkmale im Biologieunterricht vor. In diesem Beitrag wird ein Kurzfragebogen vorgestellt, der anhand inhaltlicher Selektionskriterien aus einem längeren Fragebogen (URHAHNE, MARSCH, WILDE & KRÜGER, 2011) zur Analyse konstruktivistischer Prozessmerkmale nach REINMANN und MANDL (2006) hervorgegangen ist. Das Instrument bildet die Merkmale aktiv, selbstgesteuert, emotional, situativ, sozial und konstruktivistisch mit sechs Subskalen und je zwei bis drei Items ab. In zwei Untersuchungen mit 288 Schülerinnen und Schülern (SuS), die im Mittel 11.9 Jahre, und 119 SuS, die im Mittel 16.2 Jahre alt waren, wird die Reliabilität und Validität durch probabilistische und klassische Testverfahren überprüft. Konfirmatorische Faktorenanalysen belegen, dass die Skalen theoriegemäß ein Sechs-Faktorenmodell ergeben, das alternativen Modellen überlegen ist. Raschanalysen zeigen zudem, dass die einzelnen Skalen jeweils eine Facette eines Konstrukts abbilden und*

sich die Skalierung über alle Skalen hinweg als äquidistant erweist. Ein Vergleich der Ergebnisse der Kurzskala mit denen der Langversion dokumentiert hinreichende Konstruktäquivalenz. Der Kurzfragebogen stellt damit eine zuverlässige und zeitökonomische Möglichkeit zur Erhebung konstruktivistischer Prozessmerkmale im Biologieunterricht dar.

### **Abstract**

*In biology education, constructivist learning environments are becoming increasingly important. However, there is no time-effective method for measuring characteristics of constructivist processes in biology classes. In this paper, a short questionnaire for the analysis of constructivist process characteristics according to REINMANN and MANDL (2006) based on a longer questionnaire (URHAHNE, MARSCH, WILD & KRÜGER, 2011) will be presented. Items for the short version were selected using content criteria. The instrument covers the features active, self-regulated, emotional, situational, social, and constructive with six subscales and two or three items each. In two studies with 288 students aged 11.9 years and 119 students aged 16.2 years, the measurement reliability and validity is demonstrated by probabilistic and classical test methods. Confirmatory factor analyses show that a six-factor model is superior to alternative models. Rasch analyses also reveal that the individual scales each represent one facet of a construct and that the scaling on all scales proves to be equidistant. A comparison of the results of the short scale with those of the long version documents adequate construct equivalence. The short questionnaire thus proves to be a reliable and time-effective means for assessing constructivist process characteristics in biology classes.*

## **1 Einleitung**

Konstruktivistische Lernumgebungen gelten als geeignete unterrichtliche Maßnahmen zur Vermeidung trägen Wissens (RENKL, 1996). Doch sollte überprüft werden können, inwieweit die Umsetzung konstruktivistisch orientierten Unterrichtens durch die Lehrkraft gelingt. Diese empirische Kontrolle ist durch prozessbezogene Fragebögen für Schülerinnen und Schüler (SuS), die konstruktivistische Unterrichtsmerkmale im Detail abbilden, möglich. Der Fragebogen zu den Prozessmerkmalen des gemäßigten Konstruktivismus (PgK) nach REINMANN und MANDL (2006) von URHAHNE, MARSCH, WILDE und KRÜGER (2011) ermöglicht es, den Grad der Erfüllung der Prozessmerkmale im Biologieunterricht einzuschätzen, und wurde bereits in verschiedenen empirischen Studien erfolgreich eingesetzt (bspw. MEYER-AHRENS ET AL., 2010; MEYER, BALSTER, BIRKHÖLZER & WILDE, 2011).

Wie viele andere Fragebögen ist der PgK mit 30 Items für den Einsatz im schulischen Regelbetrieb recht lang, v.a. als ein Instrument unter mehreren. Daher sollte das Beantworten von redundanten (nicht subjektiv als redundant empfundenen) Items, die die Testgüte nicht wesentlich bestimmen, vermieden

werden (vgl. BURISCH, 1997). Aus diesem Grund wird ein Kurzfragebogen zur Testung der *Prozessmerkmale des gemäßigten Konstruktivismus (Kurz-PgK)* vorgestellt. Um wiederholte Belege für die Konstruktvalidität (SMITH, 2005) zu erhalten, wird die gekürzte Version in zwei unabhängigen Stichproben auf eine vergleichbar gute Reliabilität wie die Langversion und ihre theoriegemäß sechsdimensionale Struktur hin überprüft.

## 2 Theorie

Konstruktivistische Ansätze haben einen lerntheoretischen Paradigmenwechsel geprägt (SIEBERT, 2005) und eine enorme Bedeutung für den naturwissenschaftlichen Unterricht erlangt (MATTHEWS, 1998; TOBIN, 1993). In der naturwissenschaftlichen Fachdidaktik werden konstruktivistisch orientierte Lernprozesse als bedeutsam erachtet (DUI & TREAGUST, 1998; MAROHN, 2008; MARSCH, HARTWIG & KRÜGER, 2009; WIDODO & DUIT, 2004), weil sie eine individuelle und aktive Wissenskonstruktion fördern und träges Wissen vermeiden helfen (RENKL, 1996).

REINMANN und MANDL (2006) schlagen eine gemäßigt konstruktivistische Perspektive vor, die Lernern Raum lässt, aktiv und ohne Vorschreibungen ein subjektiv geprägtes Sachverständnis aufzubauen. Gleichzeitig bieten entsprechende Lernumgebungen Struktur und Hilfe, die gerade schwächeren SuS Orientierung ermöglicht. Lernen ist dabei durch sechs Prozessmerkmale (PgK) charakterisiert und wird verstanden als ein aktiver, selbstgesteuerter, emotionaler, situativer, sozialer und konstruktiver Prozess (REINMANN & MANDL, 2006). Reinmann und Mandl haben die PgK bereits in den 90er Jahren aus der Instruktions- (DECORTE, 1993) bzw. Kognitionspsychologie (SHUELL, 1986) übernommen und als Grundlage für die Förderung von Wissenserwerb verstanden (vgl. REINMANN-ROTHMEIER & MANDL, 1998). Lernen ist ein *aktiver* Konstruktionsprozess, durch den bedeutungshaltiges Wissen erworben wird (RENKL, 1996). Dazu muss sich der Lerner konzentriert (BERG & IMHOF, 2006), bewusst und zielgerichtet (REINMANN & MANDL, 2006) mit dem Lerngegenstand auseinandersetzen. Die innere Beteiligung setzt motivierte Lerner (vgl. DECI & RYAN, 1993) voraus, die zumindest über situationales Interesse am Lerngegenstand verfügen (vgl. KRAPP, 1992; KRAPP & PRENZEL, 2011). *Selbstgesteuertes* Lernen bezieht sich sowohl auf die tatsächlichen Handlungsmöglichkeiten (WEINERT, 1982) als auch auf die Wahrnehmung von Selbststeuerung bzw. Kontrolle durch den Lerner (vgl. DECI & RYAN, 1993, 2002). Schülerreaktionen auf Lernanforderungen werden maßgeblich durch *Emotionen* mitbestimmt. Diese können auf die Anstrengungsbereitschaft der Lerner fördernd (z.B. Freu-

de, Hoffnung, Stolz etc.) oder hemmend (z.B. Langeweile, Ärger, Angst, Enttäuschung etc.) einwirken (FRENZEL, GÖTZ & PEKRUN, 2009). Die spezifischen Kontexte der *Lernsituation* bieten für Lerninhalte einen Interpretationshintergrund (vgl. MANDL, GRUBER & RENKL, 1997), der bedeutungsvolles Lernen beispielsweise durch eine Anbindung an die Lebenswelt der Lerner begünstigen kann (REINMANN & MANDL, 2006). Lernen wird als *sozial* vermittelter Prozess verstanden, der in gemeinsamer Wissenskonstruktion die Möglichkeit bietet, neue Lernerfahrungen zu machen (LAVE & WENGER, 1991). In einer sozialen Lernumgebung können sich die Lerner gegenseitig unterstützen und so gemeinsam höhere kognitive Leistungen erreichen, als sie alleine dazu in der Lage wären (vgl. WYGOTSKI, 1974). SuS beginnen ihr Lernen nicht als unbeschriebenes Blatt. Sie haben bereits Vorwissen und Erfahrungen, die es ihnen ermöglichen, sich in der Welt mit ihrer *Konstruktion* von Wirklichkeit zurechtzufinden. In Lernprozessen erfolgen eine Interpretation von neuen Informationen und die Verknüpfung mit Vorwissen (HASSELHORN & GOLD, 2006). Lernern obliegt es, welche Informationen sie aufnehmen und wie sie diese ihrem Wissen anpassen (vgl. KRÜGER, 2007). REINMANN-ROTHMEIER und MANDL (1998) nehmen an, dass die Prozessmerkmale nicht unabhängig voneinander sind, so dass für die Messung der PgK zu untersuchen bleibt, ob es sich um ein sechs- oder im Extremfall eindimensionales Konstrukt handelt.

URHAHNE ET AL. (2011) entwickelten deduktiv das erste Messinstrument zu den PgK und überprüften es mit Verfahren der klassischen Testtheorie hinsichtlich seiner Reliabilität, Faktorenstruktur und erster Validitätshinweise. Obwohl sie trotz ihres deduktiven Vorgehens bereits zufriedenstellende Ergebnisse erzielen konnten, lassen sich die Reliabilitäten und die Faktorenstruktur optimieren. Zunächst sind die internen Konsistenzen im Bereich von .69 bis .89, was auf eine eher hohe inhaltliche Breite der Items hindeutet. Des Weiteren konnten nur vier der sechs Dimensionen in der Faktorenanalyse klar voneinander getrennt werden. Zwei Items der Subskala *aktiv* wiesen substantielle Nebenladungen auf die Subskala *emotional* auf. Zwei Items der Subskala *konstruktiv* mussten aus der Faktorenanalyse entfernt werden; ein weiteres Item lud substantiell auch auf die Subskala *selbstgesteuert*. Das Merkmal *emotional* wurde erst 2006 in die Liste der PgK aufgenommen (vgl. REINMANN-ROTHMEIER & MANDL, 2001 mit fünf Prozessmerkmalen) und die fehlende Trennbarkeit der Merkmale *emotional* und *aktiv* muss daher besonders berücksichtigt werden. Die Erhebung der PgK erfolgt mit einer fünfstufigen verbalisierten Ratingskala mit Mittelkategorie (vgl. PORST, 2008, S. 81f.) der „trifft-zu-Reihe“ (ROHRMANN, 1978), für die die Annahme der Äquidistanz nicht belegt ist (RENNER, HEYDASCH & STRÖHLEIN, 2012, S. 53).

### 3 Fragestellungen

Zur Verbesserung der Erhebungsökonomie (vgl. BURISCH, 1997) und der Faktorenstruktur sollen die sechs Prozessmerkmale des gemäßigten Konstruktivismus (PgK) durch sechs Kurzskalen erfasst werden, die nach inhaltlichen Auswahlkriterien aus der Langversion der PgK (URHAHNE ET AL., 2011) entwickelt wurden. Folgende Fragen sollen in dieser Darstellung beantwortet werden: 1.) *Sind Reliabilität, Faktorenstruktur und Skalierung des Kurzfragebogens zufriedenstellend?* Zu diesem Zweck werden Kennwerte der klassischen Testtheorie, konfirmatorische Faktorenanalysen und Modelle der Item-Response-Theory (IRT) herangezogen. 2.) *Bildet die Kurzversion die Langversion des Tests hinreichend ab?* Zu diesem Zweck wurden Lang- und Kurzversion der PgK korreliert.

### 4 Hypothesen

Folgende Hypothesen werden zur Beantwortung der Forschungsfragen anhand zweier unabhängiger Stichproben überprüft:

1. Die sechs theoretisch angenommenen Faktoren konstruktivistischer Lernumgebungen lassen sich durch den Kurzfragebogen empirisch abbilden.
2. Ein sechsdimensionales Messmodell ist alternativen Messmodellen, insbesondere einem eindimensionalen Generalfaktormodell, in Bezug auf die Modellpassung empirisch überlegen.
3. Die Subskalen erfassen jeweils eine spezifische Facette gemäßigt konstruktivistischer Lernumgebungen und bilden damit im Sinne der IRT eindimensionale Facetten ab.
4. Die Abstände zwischen den fünf Messpunkten der Ratingskala sind im Sinne eines Ratingskalenmodells der IRT über die Items hinweg konstant.

In Studie 1 wurde außerdem die folgende Hypothese überprüft:

5. Kurz- und Langversion des Fragebogens korrelieren mittel bis stark miteinander, was für die Validität des Kurzfragebogens spricht.

### 5 Methodik

*Testinstrument.* Grundlage für die Entwicklung der Kurzskalen ist das 30 Items umfassende Messinstrument PgK von URHAHNE ET AL. (2011). Der Kurz-PgK umfasst 17 Items, die sich auf eine unmittelbar vorangegangene Unterrichtsstunde beziehen. Alle Formulierungen sind klar und einfach gehalten und

kommen ohne Verneinungen, Invertierungen oder inhaltliche Dopplungen aus (s. Tabelle 4). Bei der Konstruktion der Subskalen des Kurz-PgK wurde versucht, den inhaltlichen Kern des jeweiligen Theoriekonstrukts präziser zu treffen. Für das Prozessmerkmal *aktiv* ist das zentrale inhaltliche Element eine bewusst-intentionale und aktive Auseinandersetzung mit dem Lernstoff. Das Merkmal *selbstgesteuert* wird durch die Wahrnehmung von Entscheidungsmöglichkeiten repräsentiert. *Emotional* wird durch das Fragen nach positiven Emotionen beim Lernen abgebildet. Für *situativ* liegt der Kern des gemessenen Konstrukts in der Anbindung an die Lebenswelt der Lerner. Für das Prozessmerkmal *sozial* wird der Fokus auf gemeinsames Lernen und eine Wissenskonstruktion in gegenseitigem Austausch gelegt. Das Merkmal *konstruktiv* wird in seinem zentralen Inhalt durch die Berücksichtigung von Vorwissensstrukturen beschrieben. Jedes Prozessmerkmal ist in Form einer Subskala mit jeweils drei Items (bzw. zwei Items für das Merkmal *selbstgesteuert*, da ein Item aufgrund seiner psychometrischen Eigenschaften ausgeschlossen werden musste) repräsentiert (s. Tabelle 4). Im Fragebogen werden die Items in zufallsdurchmischter Reihenfolge dargeboten und von den SuS über eine fünfstufige Ratingskala (0 - trifft nicht zu, 1 - trifft kaum zu, 2 - trifft teilweise zu, 3 - trifft überwiegend zu, 4 - trifft voll zu) bewertet.

*Durchführung.* Die Entwicklung und Prüfung der Güte des Kurz-PgK erfolgt auf Basis der Studie von URHAHNE ET AL. (2011) (Studie 1) und einer zweiten Studie mit älteren SuS (Studie 2). Die Stichprobe der *Studie 1* setzte sich aus 288 SuS der Sekundarstufe I von drei niedersächsischen Gymnasien zusammen (52% Mädchen). Die SuS waren zwischen 10 und 14 Jahren alt ( $M = 11.9$ ,  $SD = 1.04$ ) und besuchten die fünfte, sechste und siebte Jahrgangsstufe (vgl. URHAHNE ET AL., 2011). Alle Gruppen wurden hierbei regulär von ihrer gewohnten Biologielehrkraft unterrichtet. Die Langversion des PgK wurde am Ende der jeweiligen Unterrichtsstunde an die Untersuchungsteilnehmer ausgeteilt. Anhand dieser Stichprobe kann daher die Hypothese 5 beantwortet werden. An *Studie 2* nahmen 119 SuS eines Gymnasiums in Nordrhein-Westfalen teil (60% Mädchen). Die SuS waren zwischen 14 und 20 Jahren alt ( $M = 16.2$ ,  $SD = 1.7$ ). Es nahmen SuS aus der neunten, elften sowie 13. Jahrgangsstufe teil. Der Kurz-PgK wurde im Rahmen eines Schulpraktikums für Lehramtsstudierende der Universität Bielefeld im Fach Biologie eingesetzt. Auch hier wurde das Messinstrument am Ende der Unterrichtsstunden ausgeteilt.

*Statistische Auswertung.* Nach der einführenden Darstellung deskriptiver Statistiken erfolgt die Analyse der dimensional Struktur des Messinstrumentes mittels konfirmatorischer Faktorenanalysen (KFAs). Als Schätzer kommt

dabei das Maximum-Likelihood-Verfahren zur Anwendung. Die Bewertung der Datenpassung auf die theoretischen Modelle erfolgt über verschiedene Fit-Statistiken (HU & BENTLER, 1995). Weil der  $\chi^2$ -Wert stark von der Stichprobengröße abhängig ist und häufig signifikant ausfällt, werden zusätzlich der Root Mean Square Error of Approximation (RMSEA) mit einem günstigen Wert ab  $RMSEA \leq .06$ , der Comparative Fit Index (CFI) mit einem günstigen Wert ab  $\geq .95$  sowie der Tucker Lewis Index (TLI) mit einem günstigen Wert ab  $\geq .95$  (ULLMAN, 2007) berichtet. Neben diesen komparativen Fit-Statistiken werden die Informationskriterien Akaike's Information Criterion (AIC) und Bayesian Information Criterion (BIC) zur Überprüfung der Anpassungsgüte herangezogen. Für AIC und BIC existieren keine absoluten Werte, die als günstig angesehen werden; vielmehr erfolgt ein relativer Vergleich von AIC und BIC zwischen konkurrierenden Modellen. Ein geringer AIC oder BIC spricht dabei für eine bessere Modellgüte. Wenn bei der Bewertung des Modellfits die unterschiedlichen Fit-Statistiken nicht immer gleiche Schlussfolgerungen nahe legen, empfiehlt sich die Betrachtung des Gesamtmusters der vorliegenden Indices und eine relativ stärkere Gewichtung des RMSEA, der nachweislich die günstigsten Eigenschaften aufweist (vgl. HU & BENTLER, 1995; YU, 2002). Aussagen zur Messgenauigkeit der Subskalen des im Rahmen einer KFA als am besten identifizierten Modells werden anhand von Cronbachs Alpha als untere Schranke der Reliabilität getätigt. Zusätzlich zur klassischen Testtheorie, die lineare Zusammenhänge zwischen manifesten und latenten Variablen annimmt, erfolgt die Analyse der psychometrischen Eigenschaften des Messkonstrukts mittels IRT-Modellen. Diese beinhalten ein theoretisches Messmodell mit dezidierten Annahmen über den Zusammenhang zwischen manifesten Itemantworten und latentem Konstrukt, die empirisch überprüfbar sind (EMBRETSON & REISE, 2000) und zur Überprüfung der Eindimensionalität herangezogen werden können. IRT-Modelle gehen dabei über die Annahmen von KFAs hinaus und beinhalten eine strengere Überprüfung der psychometrischen Güte einzelner Skalen (WILSON, 2005). KFAs werden in MPlus 5 (MUTHÉN & MUTHÉN, 2007) und IRT-Modelle in ConQuest 3.1 (WU, ADAMS & HALDANE, 2005) berechnet.

## 6 Ergebnisse und Diskussion

Skalenkennwerte und die latenten Skaleninterkorrelationen sind für Studie 1 in Tabelle 1 und für Studie 2 in Tabelle 2 dargestellt. Um die Reliabilität des Kurz-PgK zu überprüfen, wurden Cronbachs Alpha-Koeffizienten für die einzelnen Subskalen berechnet und in den Diagonalen von Tabelle 1 bzw. Tabelle

2 abgebildet. Beim Vergleich beider Studien fällt eine hohe absolute sowie relative Konsistenz der Ergebnisse zwischen den Stichproben auf. Die von 0 bis 4 gestufte Ratingskala wird in beiden Stichproben in ihrer ganzen Breite genutzt. Es treten substantielle interindividuelle Unterschiede auf. Hinweise auf bedeutsame Boden- oder Deckeneffekte gibt es nicht.

**Tabelle 1:** Interne Konsistenzen, latente Korrelationen und Skalenkennwerte der 6 Dimensionen in Studie 1 (n = 288).

	aktiv	selbst- gesteuert	emotional	situativ	sozial	konstruktiv
aktiv	<b>.77</b>					
selbstgesteuert	.43 **	<b>.70</b>				
emotional	.75 **	.37 **	<b>.81</b>			
situativ	.21 **	.49 **	.34 **	<b>.82</b>		
sozial	.37 **	.19 **	.40 **	.16 *	<b>.79</b>	
konstruktiv	.44 **	.33 **	.39 **	.43 **	.16 *	<b>.69</b>
Mittelwert	2.54	2.26	2.62	2.60	2.30	1.52
SW	.79	1.07	1.05	.84	1.17	1.06
ICC (1)	.59 **	.60 **	.55 **	.53 **	.43 **	.50 **

Anmerkungen: Cronbachs Alphas der einzelnen Skalen in der Hauptdiagonalen.

\*p < .05; \*\*p < .01.

Cronbachs Alpha liegt für die einzelnen Subskalen im Bereich von .68 bis .89, so dass sich alle Dimensionen, trotz geringer Itemanzahl, als hinreichend messgenau erweisen. Die etwas geringere Messgenauigkeit der Skala *selbstgesteuert* ist v.a. auf die Zahl von nur zwei Items zurückzuführen. Weiterhin sind in den Tabellen 1 und 2 die Intraklassenkorrelationen abgebildet. Sie beschreiben die Genauigkeit, mit der ein Schüler im Durchschnitt das entsprechende Unterrichtsmerkmal beurteilt (LÜDTKE, TRAUTWEIN, KUNTER & BAUMERT, 2006). Die ICC(1)-Werte deuten auf eine hinreichende Messzuverlässigkeit hin. Die latenten Korrelationen zwischen den angenommen Dimensionen liegen bis auf einige Korrelationen mit dem Merkmal *emotional* im moderaten Bereich. Sie bestätigten zunächst das angenommene Bild der relativen Unabhängigkeit der einzelnen Dimensionen, die zwar einerseits Überlappungen haben (vgl. REINMANN-ROTHMEIER & MANDL, 1998), zugleich aber auch substantielle Eigenständigkeit aufweisen. Lediglich die beiden Skalen *aktiv* und *emotional* zeigen in beiden Stichproben recht hohe Korrelationen (vgl. URHAHNE ET AL., 2011), während die anderen Subskalen in der Mehrheit lediglich 10% bis 25% Varianz miteinander teilen.

**Tabelle 2:** Interne Konsistenzen, latente Korrelationen und Skalenkennwerte der 6 Dimensionen in Studie 2 (n = 119).

	aktiv	selbst- gesteuert	emotional	situativ	sozial	konstruktiv
aktiv	<b>.84</b>					
selbstgesteuert	.40 **	<b>.72</b>				
emotional	.66 **	.60 **	<b>.86</b>			
situativ	.44 **	.44 **	.52 **	<b>.89</b>		
sozial	.35 **	.34 **	.33 **	.27 **	<b>.68</b>	
konstruktiv	.29 **	.15	.39 **	.18 *	.33 **	<b>.74</b>
Mittelwert	2.88	2.85	3.42	2.99	2.48	1.76
SW	.69	.76	.58	.68	.97	.83
ICC (1)	.72 **	.64 **	.68 **	.51 **	.47 **	.57 **

Anmerkungen: Cronbachs Alphas der einzelnen Skalen in der Hauptdiagonalen.

\* $p < .05$ ; \*\* $p < .01$ .

Die Darstellung der latenten Skaleninterkorrelationen in den Tabellen 1 und 2 setzt implizit voraus, dass eine Trennung aller sechs theoretisch postulierten Dimensionen auch empirisch gegeben ist. Diese Fragestellung wurde über konfirmatorische Faktorenanalysen überprüft. Das für den Kurz-PgK in Hypothese 1 angenommene sechsdimensionale Modell spiegelt sich in beiden Studien in den Daten wider und zeigt einen guten bis akzeptablen Modellfit über alle Fitindices hinweg (s. Tabelle 3). Sowohl der absolute Vergleich zwischen der Passung eines sechsdimensionalen Modells und eines eindimensionalen Modells mit einem übergeordneten Faktor als auch die zugehörigen  $\chi^2$ -Differenzentests ( $\chi^2 = 732.25$ ,  $df = 14$ ,  $p < .001$  in Studie 1;  $\chi^2 = 895.70$ ,  $df = 32$ ,  $p < .001$  in Studie 2) sprechen in beiden Stichproben deutlich für das sechsdimensionale Modell. Basierend auf den hohen Korrelationen zwischen den Dimensionen *aktiv* und *emotional* in beiden Stichproben (vgl. URHAHNE ET AL. 2011) wurde weiterhin ein post-hoc abgeleitetes fünfdimensionales Modell getestet, das die Dimensionen *aktiv* und *emotional* miteinander kombiniert (s. Tabelle 3). Gegenüber dem ursprünglichen sechsdimensionalen Modell führt die Zusammenlegung der Dimensionen *aktiv* und *emotional* zu einem Absinken der Modellpassung ( $\chi^2$ -Differenzentest:  $\chi^2 = 67.21$ ,  $df = 5$ ,  $p < .001$  in Studie 1;  $\chi^2$ -Differenzentest:  $\chi^2 = 72.56$ ,  $df = 5$ ,  $p < .001$  in Studie 2), so dass auch das post-hoc abgeleitete fünfdimensionale Alternativmodell die Daten weniger gut beschreibt als das sechsdimensionale Modell.

**Tabelle 3:** Modellfit für das 6-, das 1- und das 5-dimensionale Modell in Studie 1 (n = 288) und Studie 2 (n = 119).

		$\chi^2$	df	p	CFI	TLI	RMSEA	KI <sub>RMSEA</sub> %	AIC	BIC
(1)	6 Dimensionen (Studie 1)	189.71	104	< .001	.95	.94	.053	.041-.065	13970	14212
(2)	1 Dimension (Studie 1)	921.96	119	< .001	.53	.46	.153	.144-.162	14672	14859
(3)	5 Dimensionen (Studie 1)	256.92	109	< .001	.91	.89	.069	.058-.080	14027	14251
(4)	6 Dimensionen (Studie 2)	155.07	104	< .001	.95	.93	.064	.042-.085	4321	4505
(5)	1 Dimension (Studie 2)	1050.77	136	< .001	.55	.49	.171	.156-.186	4668	4648
(6)	5 Dimensionen (Studie 2)	227.63	109	< .001	.87	.84	.096	.078-.113	4384	4554

Anmerkungen:  $\chi^2$  = empirischer  $\chi^2$ -Wert; df = Freiheitsgrade; p = Überschreitungswahrscheinlichkeit; CFI = Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA = Root Mean Square Error of Approximation; KI<sub>RMSEA</sub> = 90%iges Konfidenzintervall für den RMSEA; AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion.

Wie oben bereits festgestellt, beschreiben REINMANN-ROTHMEIER und MANDL (2001) zunächst nur fünf Prozessmerkmale – das Merkmal *emotional* wurde erst im revidierten Text ergänzt (REINMANN & MANDL, 2006). Die hohen Korrelationen zwischen *aktiv* und *emotional* lassen sich damit jedoch nicht erklären, denn das Prozessmerkmal *emotional* kann nicht als Ausschärfung eines Teilaspektes des Prozessmerkmals *aktiv* betrachtet werden. Beide lassen sich konzeptuell gut voneinander abgrenzen. Auch die Operationalisierungen (*aktiv* als tätigkeitsbezogene Auseinandersetzung mit dem Lerngegenstand, *emotional* als positive Erlebnisqualitäten während der Tätigkeit) im Kurz-PgK sind konzeptuell distinkt. Eine aktive Beteiligung und positive Gefühle während der Tätigkeit traten jedoch anscheinend oft gemeinsam auf. Die Trennung der beiden Dimensionen ließ sich empirisch dennoch gut belegen, denn die internen Konsistenzen sind in beiden Studien gut und die Faktoren laden jeweils hoch und spezifisch auf der theoretisch angenommenen Subskala. Insbesondere spricht auch der Vergleich des sechsdimensionalen Messmodells mit dem fünfdimensionalen für das differenzierte Modell. Die Faktorladungen für das sechsdimensionale Modell sind alle substanziell und die Kommunalitäten liegen mit wenigen Ausnahmen über dem von HAIR, ANDERSON, TATHAM und BLACK (1998) identifizierten Idealbereich von  $> .40$  (s. Tabelle 4). Die einzelnen Items laden ausschließlich auf den theoriegemäß angenommenen Faktoren und das Zulassen von Nebenladungen verbessert den Modellfit nicht wesentlich.

**Tabelle 4:** Standardisierte Faktorladung, Kommunalitäten und MNSQ-Werte für die 6 Dimensionen in Studie 1 (n = 288) und Studie 2 (n = 119).

Itemformulierung: In dieser Unterrichtsstunde...	stand. Ladung	Kommu- nalität	MNSQ
<b>Skala 1: aktiv</b>			
...war ich beim Lernen tätig.	.69** / .77**	.48 / .59	.86 / 1.05
...war ich beim Lernen eifrig.	.70** / .91**	.49 / .83	.94 / .95
...war ich beim Lernen aktiv.	.79** / .74**	.62 / .55	1.14 / 1.27
<b>Skala 2: selbstgesteuert</b>			
...konnte ich selbst bestimmen, was ich lerne.	.58** / .78**	.34 / .61	1.02 / 1.07
...konnte ich selbst entscheiden, womit ich lerne.	.86** / .73**	.74 / .53	.95 / .89
<b>Skala 3: emotional</b>			
...hatte ich Lust zu lernen.	.79** / .61**	.62 / .37	.98 / 1.12
...hatte ich Spaß beim Lernen.	.78** / .89**	.61 / .79	1.09 / 1.02
...habe ich mich beim Lernen wohlgeföhlt.	.74** / .65**	.55 / .42	.90 / .99
<b>Skala 4: situativ</b>			
...habe ich etwas gelernt, was ich im Alltag nutzen kann.	.71** / .62**	.50 / .38	1.16 / 1.21
...habe ich etwas gelernt, was ich gut gebrauchen kann.	.85** / .79**	.72 / .62	1.04 / .91
...habe ich etwas gelernt, was mir im Leben weiterhilft.	.77** / .66**	.59 / .44	1.08 / .73
<b>Skala 5: sozial</b>			
...habe ich beim Lernen mit anderen zusammen gearbeitet.	.81** / .84**	.66 / .71	1.05 / .97
...habe ich mich beim Lernen mit anderen ausgetauscht.	.74** / .87**	.55 / .76	.95 / .86
...habe ich mit anderen zusammen gelernt.	.68** / .76**	.46 / .58	1.12 / 1.10
<b>Skala 6: konstruktiv</b>			
...habe ich auf vorhandenes Wissen aufgebaut.	.72** / .77**	.52 / .59	.91 / 1.12
...habe ich auf vorhandene Kenntnisse zurückgegriffen.	.68** / .94**	.46 / .88	.94 / 1.12
...habe ich neues Wissen mit altem Wissen verknüpft.	.57** / .84**	.32 / .71	1.24 / .99

Anmerkungen: Darstellung erfolgt nach Studie 1 / Studie 2; stand. = standardisiert. E(MNSQ) = 1. Werte kleiner als 1 indizieren in IRT-Modellen zu hohe Trennschärfen, Werte größer als 1 zu niedrige.

\*\*p < .01.

Die bisher präsentierten Ergebnisse zeigen in beiden Stichproben, dass sich die theoretisch angenommenen Faktoren auch als Kurzfassung empirisch abbilden lassen (Hypothese 1) und dass ein sechsdimensionales Modell alternativen Modellen überlegen ist (Hypothese 2).

Ergänzend zu den KFAs wurden zur Klärung der Hypothese 3 für die sechs Dimensionen getrennte Skalierungen im Rahmen der IRT vorgenommen, die im Vergleich zum bislang gewählten Vorgehen striktere Annahmen über die Beziehung zwischen manifesten Itemantworten und latenter Variable machen und dabei überprüfbare Annahmen über die Dimensionalität des zugrundeliegenden Konstrukts postulieren (BOND & FOX, 2001). Zur Überprüfung des Modellfits kommen in der IRT neben dem globalen Modellfit v.a. Mean Square (MNSQ)-Statistiken als Itemfitmaße (WILSON, 2005) zur Anwendung. Die MNSQ-Werte liegen für ein einparametrisches Modell mit gestuften Antworten (d.h. Partial Credit Model, PCM, mit den fünf Antwortkategorien pro Item) in beiden Stichproben durchgängig im von BOND und FOX (2001) genannten Tole-

ranzbereich von .75 bis 1.33 für alle Items einer Subdimension, so dass die einzelnen Dimensionen des gemäßigten Konstruktivismus als eindimensionale Facetteninstrumente im Sinne von IRT-Skalierungen verstanden werden können. Das in Hypothese 3 applizierte PCM lässt allerdings variierende Kategorienabstände zwischen Items zu und beschreibt damit ein verhältnismäßig liberales Modell (BOND & FOX, 2001). Das strengere Rating Scale Model (RSM; EMBRETSON & REISE, 2000) nimmt an, dass zusätzlich zu den gängigen Annahmen von IRT-Modellen (vgl. WILSON, 2005) die Abstände zwischen den Antwortkategorien über alle Items hinweg identisch sind, also dass die Antwortskala bei allen Items identisch funktioniert. Diese zusätzliche Restriktion im Rahmen des RSM, die in Hypothese 4 formuliert ist, kann von den Daten angemessen erfüllt werden: Erstens fällt ein  $\chi^2$ -Differenzentest zwischen dem PCM und dem sparsameren RSM (Abstände zwischen den Antwortkategorien werden für alle Items gemeinsam und nicht separat für jedes Item geschätzt) für alle sechs Subdimensionen nicht signifikant aus ( $\chi^2 = 1.20$  bis  $8.27$ ;  $df = 6$ ;  $p = ns$ ). Zweitens liegen die MNSQ-Werte für das RSM ebenfalls noch innerhalb der angegebenen Grenzwerte von .75 bis 1.33 (s. Tabelle 4). Diese Befunde zeigen sich sowohl in Studie 1 als auch in Studie 2. Es kann also tatsächlich jede Subskala als *eine* latente Variable aufgefasst werden. Das Instrument verfügt also auch bezüglich der IRT-Modelle über gute psychometrische Eigenschaften. Abschließend wurde in Hypothese 5 überprüft, ob der Kurz-PgK hinreichend mit der ausführlicheren Langversion korreliert. Die Korrelationen zwischen den Subskalen des Kurz-PgK und der Langversion liegen zwischen .87 und .97 (alle  $p < .01$ ), so dass von einer hohen Konstruktüberlappung sowie von hinreichender Konstruktäquivalenz ausgegangen werden kann.

## 7 Fazit

Es ließen sich sowohl in einer Stichprobe von SuS der Sekundarstufe I (Studie 1), als auch einer weiteren mit dem Schwerpunkt in der Sekundarstufe II (Studie 2) gute und nahezu identische psychometrische Eigenschaften des Kurz-PgK nachweisen. Zusammenfassend stützen die Analysen eine sechsdimensionale Struktur im Rahmen konfirmatorischer Faktorenanalysen (Hypothese 1) und zeigen deren Überlegenheit gegenüber zwei alternativen Messmodellen (Hypothese 2). Die insgesamt gute Passung bestätigt sich ebenso in den zusätzlich durchgeführten IRT-Analysen (Hypothesen 3 und 4). Die Korrelation mit der Langversion ist hoch (Hypothese 5).

Der Kurz-PgK konnte damit den Anforderungen an ein objektives, reliables und valides Messinstrument im Rahmen der klassischen und der

probabilistischen Testtheorie in beiden Stichproben genügen. Dieser Befund spricht für eine Generalisierbarkeit des Konstrukts. Die Eignung in den beiden Stichproben unterschiedlichen Alters (vgl. SMITH, 2005) verweist zudem darauf, dass der Kurz-PgK für Langzeitstudien geeignet ist, zumal diese zeitökonomische Version des Fragebogens bei geringem Informationsverlust im Vergleich zur Langversion des PgK (vgl. BURISCH, 1997) eine Verwendung im Regelunterricht erheblich einfacher macht. Die Skala *selbstgesteuert* könnte ungeachtet der vorliegenden psychometrischen Güte in zukünftigen Untersuchungen durch ein neues Item ergänzt werden, um auch in dieser Skala drei Items erheben zu können. Wünschenswert wären zudem weitere Einsätze des Kurz-PgK in größeren Stichproben als die der Studie 2 (vgl. SMITH, 2005).

Für viele Konstrukte, die in der Fachdidaktik von Interesse sind, könnte eine prozessbezogene Kontrolle des Unterrichts anhand der PgK-Skalen bei vertretbarem Aufwand wertvolle Zusatzinformationen liefern. Laut REINMANN und MANDL (2006) korrespondiert konstruktivistisch orientierter Unterricht (operationalisiert durch die sechs Prozessmerkmale) mit guten Schülerleistungen und guter Unterrichtsqualität. Die Einsatzmöglichkeiten des Kurz-PgK sind somit breit gefächert. Beispielsweise wäre eine Verknüpfung mit Leistungsüberprüfungen möglich und würde neben einer Darstellung des Status quo Informationen darüber liefern, wie es zu diesem gekommen ist. Das Instrument ließe sich auch in Untersuchungen zur Unterrichtsqualität, z.B. „Merkmale der Unterrichtsgestaltung“ (z.B. HELMKE, 2004; NEUHAUS, 2007; SEIDEL & SHAVELSON, 2007), bzw. Merkmalen guten Unterrichts, z.B. „Lernförderliches Klima“ (KRÜGER, KLOSS & CUADROS, 2009; MEYER, 2004), einsetzen. Untersuchungen dieser Art würden den Prozess der Konstruktvalidierung der PgK weiter voranbringen (vgl. CRONBACH & MEEHL, 1955; SMITH, 2005). Erste quasi-experimentelle Hinweise auf Konstruktvalidität liegen für die Langversion vor (bspw. MEYER-AHRENS ET AL., 2010; MEYER ET AL., 2011).

Insgesamt kann der Kurz-PgK aus psychometrischer Sicht für eine Beschreibung des Unterrichtsprozesses bezüglich gemäßigt konstruktivistischer Prozessmerkmale empfohlen werden und seine Nützlichkeit für die fachdidaktische Forschung – aber auch für Biologielehrkräfte im eigenen Unterricht – als hoch eingeschätzt werden.

## Zitierte Literatur

- BERG, D. & IMHOF, M. (2006): Aufmerksamkeit und Konzentration. In: ROST, D. H. [Hrsg.]: Handwörterbuch Pädagogische Psychologie. PVU, Weinheim, 45-53.
- BOND, T. G. & FOX, C. M. (2001): Applying the Rasch model: Fundamental measurement in the human sciences. Erlbaum, Mahwah, NJ.

- BURISCH, M. (1997): Test length and validity revisited. *EJP* **11**, 303-315.
- CRONBACH, L. J. & MEEHL, P. E. (1955): Construct validity in psychological tests. *Psychol Bull* **52** (4), 281-302.
- DECI, E. L. & RYAN, R. M. (1993): Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Z.f.Päd.* **39** (2), 223-238.
- DECI, E. L. & RYAN, R. M. (2002): *Handbook of self-determination research*. University of Rochester, Rochester.
- DECORTE, E. (1993): Learning Theory and Instructional Science. Invited paper presented at the Final Planning Workshop of the ESF-Programme "Learning in Humans and Machines".
- DUIT, R. & TREAGUST, D. (1998): Learning in science: From behaviourism towards social constructivism and beyond. In: FRASER, B. & TOBIN, K. [Hrsg.]: *International Handbook of Science Education*. Kluwer, Dordrecht, Niederlande., 3-25.
- EMBRETSON, S. E. & REISE, S. P. (2000): *Item response theory for psychologists*. Erlbaum, Mahwah, NJ.
- FRENZEL, A. C., GÖTZ, T. & PEKRUN, R. (2009): Emotionen. In: WILD, E. & MÖLLER, J. [Hrsg.]: *Lehrbuch Pädagogische Psychologie*. Springer, Heidelberg, 205-234.
- HAIR, J. F., ANDERSON, R. E., TATHAM, R. L. & BLACK, W. (1998): *Multivariate data analysis*. Prentice Hall, Upper Saddle River, NJ.
- HASSELHORN, M. & GOLD, A. (2006): *Pädagogische Psychologie. Erfolgreiches Lernen und Lehren*. Kohlhammer, Stuttgart.
- HELMKE, A. (2004): *Unterrichtsqualität: Erfassen. Bewerten. Verbessern*. Kallmeyer, Seelze.
- HU, L., & BENTLER, P. M. (1995): Evaluating model fit. In: HOYLE, R. H. [Hrsg.]: *Structural equation modeling*. Sage, London, 76-99.
- KRAPP, A. (1992): Konzepte und Forschungsansätze zur Analyse des Zusammenhangs von Interesse, Lernen und Leistung. In: KRAPP, A. & PRENZEL, M. [Hrsg.]: *Interesse, Lernen, Leistung. Neuere Ansätze einer pädagogisch-psychologischen Interessenforschung*. Aschendorff, Münster, 9-52.
- KRAPP, A. & PRENTZEL, M. (2011): Research on interest in science: Theories, methods, and findings. *Int. J. Sci. Edu.* **33** (1), 27-50.
- KRÜGER, D. (2007): Die Conceptual Change-Theorie. In: KRÜGER, D. & VOGT, H. [Hrsg.]: *Theorien in der biologieidaktischen Forschung*. Springer, Heidelberg., 165-175.
- KRÜGER, D., KLOSS, L. & CUADROS, I. (2009): Was macht einen "guten" Biologielehrer aus? - Befragungen von Dozenten in der Didaktik der Biologie-Lehramtsstudierenden an deutschen Hochschulen. *IDB* **18**, 63-88.
- LAVE, J. & WENGER, E. (1991): *Situated learning: Legitimate peripheral participation*. Cambridge University Press, Cambridge.
- LÜDTKE, O., TRAUTWEIN, U., KUNTER, M. & BAUMERT, J. (2006): Analyse von Lernumwelten: Ansätze zur Bestimmung der Reliabilität und Übereinstimmung von Schülerwahrnehmungen. *ZfPP*. **20**, 85-96.
- MANDL, H., GRUBER, H. & RENKL, A. (1997): Situiertes Lernen in multimedialen Lernumgebungen. In: ISSING, L. & KLIMSA, P. [Hrsg.]: *Information und Lernen mit Multimedia*. Beltz, Weinheim., 167-178.
- MAROHN, A. (2008): „Choice2learn“ - eine Konzeption zur Exploration und Veränderung von Lernervorstellungen im naturwissenschaftlichen Unterricht. *ZfDN* **14**, 57-83.
- MARSCH, S., HARTWIG, C. & KRÜGER, D. (2009): Biologieunterricht. Beurteilung konstruktivistisch orientierter Lernumgebungen. *ZfDN* **15**, 109-130.
- MATTHEWS, M. R. (1998): Introductory comments on philosophy and constructivism in science Education. In: MATTHEWS, M. R. [Hrsg.]: *Constructivism in science education. A philosophical examination*. Kluwer, Dordrecht, Niederlande., 1-10.
- MEYER-AHRENS, I., BÄTZ, K., DAMERAU, K., DEMMING, A., GREWE, M., KLINGSIEK, J., SCHULZ, A., URHAHNE, D. & WILDE, M. (2010): Einfluss kollektiver Schülerentscheidungen auf die Lernerwahrnehmung der konstruktivistischen Orientierung ihres Unterrichts. *EP* **24** (2), 115-130.
- MEYER, A., BALSTER, S., BIRKHÖLZER, C. & WILDE, M. (2011): Der Einfluss von lebenden Tieren als Unterrichtsmittel auf die Lernerwahrnehmung der konstruktivistischen Orientierung ihres Biologieunterrichts. *ZfDN* **17**, 339-355.
- MEYER, H. (2004): *Was ist guter Unterricht?* Cornelsen Scriptor, Berlin.
- MUTHÉN, B. O., & MUTHÉN, L. K. (2007): *MPlus*. Muthén & Muthén, Los Angeles, CA.

- NEUHAUS, B. (2007): Unterrichtsqualität als Forschungsfeld für empirische biomedizinische Forschung. In: KRÜGER, D. & VOGT, H. [Hrsg.]: Handbuch der Theorien in der biomedizinischen Forschung. Springer, Heidelberg., 243-254.
- PORST, R. (2008): Fragebogen: Ein Arbeitsbuch. VS Verlag für Sozialwissenschaften, Wiesbaden.
- REINMANN, G. & MANDL, H. (2006): Unterrichten und Lernumgebungen gestalten. In: KRAPP, A. & WEIDEMANN, B. [Hrsg.]: Pädagogische Psychologie. Beltz, Weinheim, 613-658.
- REINMANN-ROTHMEIER, G. & MANDL, H. (2001): Unterrichten und Lernumgebungen gestalten. In: KRAPP, A. & WEIDEMANN, B. [Hrsg.]: Pädagogische Psychologie. Beltz, Weinheim, 601-646.
- REINMANN-ROTHMEIER, G. & MANDL, H. (1998): Wissensvermittlung: Ansätze zur Förderung des Wissenserwerbs. In: KLIX, F. & SPADA, H. [Hrsg.], Wissen. Enzyklopädie der Psychologie, Themenbereich C, Serie II, Band 7. Hogrefe, Göttingen, 457-500.
- RENNER, K.-H., HEYDASCH, T. & STRÖHLEIN, G. (2012): Forschungsmethoden der Psychologie. Von der Fragestellung zur Präsentation. Springer VS, Wiesbaden.
- RENKL, A. (1996): Träges Wissen: Wenn Erlerntes nicht genutzt wird. Psychol Rundsch **47**, 78-92.
- ROHRMANN, B. (1978): Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. Z. f. Sozialpsychol. **9**, 222-245.
- SEIDEL, T. & SHAVELSON, R. J. (2007): Teaching effectiveness research in the last decade: Role of theory and research design in disentangling meta-analysis results. Rev Edu Res **77** (4), 454-499.
- SHUELL, T.J. (1986): Cognitive Conceptions of Learning. Rev of Edu Res **56** (4), 411-436.
- SIEBERT, H. (2005): Pädagogischer Konstruktivismus. Lernzentrierte Pädagogik in Schule und Erwachsenenbildung. Beltz, Weinheim.
- SMITH, G. T. (2005): On construct validity. Issues of method and measurement. Psychol Assess **17** (4), 396-408.
- TOBIN, K. (1993): The practice of constructivism in science education. Erlbaum, Hillsdale, NJ.
- ULLMANN, J. B. (2007): Structural equation modeling. In: TABACHNIK, B. G. & FIDELL, L. S. [Hrsg.]: Using multivariate statistics. Pearson, Boston, MA., 676-780.
- URHAHNE, D., MARSCH, S., WILDE, M. & KRÜGER, D. (2011): Die Messung konstruktivistischer Unterrichtsmerkmale auf der Grundlage von Schülerurteilen. PEU **58**, 116-127.
- WEINERT, F. E. (1982): Selbstgesteuertes Lernen als Voraussetzung, Methode und Ziel des Unterrichts. U.wiss. **10**, 99-110.
- WIDODO, A. & DUIT, R. (2004): Konstruktivistische Sichtweisen vom Lehren und Lernen und die Praxis des Physikunterrichts. ZfDN **10**, 232-254.
- WILSON, M. (2005): Constructing measures: An item response modeling approach. Erlbaum, Mahwah, NJ.
- WU, M., ADAMS, R. & HALDANTE, S. (2005): ConQuest (Version 3.1). University of California, Berkeley, CA.
- WYGOTSKI, L. S. (1974): Denken und Sprechen. Fischer, Stuttgart.
- YU, C.-Y. (2002): Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcome. University of California, Los Angeles, CA.

