

Zehner, Fabian; Weis, Mirjam; Vogel, Freydis; Leutner, Detlev; Reiss, Kristina  
**Kollaboratives Problemlösen in PISA 2015. Deutschland im Fokus**

*formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:*

*formally and content revised edition of the original source in:*

*Zeitschrift für Erziehungswissenschaft 22 (2019) 3, S. 617-646*



Bitte verwenden Sie in der Quellenangabe folgende URN oder DOI /

Please use the following URN or DOI for reference:

urn:nbn:de:0111-pedocs-176046

10.25656/01:17604

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-176046>

<https://doi.org/10.25656/01:17604>

#### Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

#### Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

#### Kontakt / Contact:

**peDOCS**

DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation

Informationszentrum (IZ) Bildung

E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)

Internet: [www.pedocs.de](http://www.pedocs.de)

Mitglied der

  
Leibniz-Gemeinschaft

## Kollaboratives Problemlösen in PISA 2015: Deutschland im Fokus

Collaborative Problem Solving in PISA 2015: Focusing on Germany

---

Fabian Zehner<sup>1</sup>, Mirjam Weis<sup>2</sup>, Freydis Vogel<sup>3</sup>, Detlev Leutner<sup>4</sup> und Kristina Reiss<sup>2</sup>

<sup>1</sup>DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation, Zentrum für technologiebasiertes Assessment (TBA), <sup>2</sup>Technische Universität München, TUM School of Education, Zentrum für Internationale Bildungsvergleichsstudien (ZIB), <sup>3</sup>Technische Universität München, TUM School of Education, Lehrstuhl für Schulpädagogik, <sup>4</sup>Universität Duisburg-Essen, Fakultät für Bildungswissenschaften, Lehrstuhl für Lehr-Lernpsychologie

---

This is a post-peer-review, pre-copyedit version of an article published in *Zeitschrift für Erziehungswissenschaft*. The final authenticated version is available online at:

<http://dx.doi.org/10.1007/s11618-019-00874-4>

Zehner, F., Weis, M., Vogel, F., Leutner, D. & Reiss, K. (2019). Kollaboratives Problemlösen in PISA 2015: Deutschland im Fokus. *Zeitschrift für Erziehungswissenschaft*, 22(3), 617–646.

doi: 10.1007/s11618-019-00874-4

## Zusammenfassung

Dieser Beitrag fokussiert die Ergebnisse in Deutschland zum internationalen Vergleich kollaborativer Problemlösekompetenz bei Fünfzehnjährigen im *Programme for International Student Assessment* (PISA) 2015 und berichtet Ergebnisse einer Kreuzvalidierung der Skalierung. Eingesetzt wurde ein neuer computerbasierter Test, der die Schülerinnen und Schüler mit simulierten Gruppenmitgliedern Probleme lösen lässt. Daten von  $n = 124.994$  Fünfzehnjährigen aus 51 Staaten zur kollaborativen Problemlösekompetenz wurden erhoben. Die Schülerinnen und Schüler in Deutschland weisen eine überdurchschnittliche Kompetenz auf (525 Punkte), liegen eine viertel Standardabweichung unter dem OECD-Spitzenstaat Japan (552 Punkte) und eine viertel Standardabweichung über dem OECD-Schnitt (500 Punkte). In allen Staaten weisen Mädchen höhere Werte auf als Jungen. Während der Anteil hochkompetenter Jugendlicher in Deutschland vergleichbar hoch mit den Spitzenstaaten ausfällt, erreichen 21 Prozent nur Kompetenzstufe I oder bleiben darunter, doppelt so viele wie in Japan. Der Beitrag präsentiert zudem nationale Ergebnisse, liefert empirische Evidenz zur Qualität des Tests und diskutiert diesen kritisch.

*Schlagwörter:* Computerbasierte Erhebung, Kollaboration, Kollaborative Problemlösekompetenz, Problemlösen, Szenarienbasiertes Assessment

---

## Abstract

Focusing on Germany, this article presents results from the international comparison of fifteen-year-olds in collaborative problem solving in the *Programme for International Student Assessment* (PISA) 2015 as well from a cross validation of the scaling. A new computer-based test was used requesting students to solve a problem jointly with simulated group members. Data from  $n = 124,994$  fifteen-year-olds in 51 countries were assessed in collaborative problem solving. The German mean competence level (525 points) is a quarter standard deviation above the OECD average (500 points) and a quarter standard deviation below the OECD's top performing country Japan (552 points). In all participating countries, girls outperform boys. While the percentage of top-performing students in Germany is comparable to proportions in the best-performing OECD countries, 21 percent of the students in Germany only reach competence level I or below, the rate being twice as high as in Japan. Further national results are presented as well as empirical evidence on the quality of the test, which is critically discussed.

*Keywords:* Computer-Based Assessment, Collaboration, Collaborative Problem Solving, Problem Solving, Scenario-Based Assessment

### **Kollaboratives Problemlösen in PISA 2015: Deutschland im Fokus**

Sowohl in der Schule als auch im Arbeits- und Privatleben werden Menschen häufig mit der Aufgabe konfrontiert, Probleme zu lösen, für welche die zur Lösung benötigten Informationen oder Fertigkeiten über verschiedene Individuen verteilt sind (Zehner, Weis, Vogel, Leutner & Reiss, 2017). In diesen Situationen ist gelingende Kollaboration nicht nur förderlich, sondern eine notwendige Voraussetzung. Speziell in der Berufsausübung sehen sich Arbeitende immer mehr nicht-routinierten und interpersonal zu lösenden Problemen gegenüber (Hermann et al., 2017, S. 243f.; Levy & Murnane, 2013, S. 18), aber auch in der Bildungspraxis findet sich eine verstärkte Nutzung kollaborativer Lernmittel (Mayrberger & Bettinger, 2014, S. 155). Daher untersucht das *Programme for International Student Assessment* (PISA), initiiert durch die *Organisation for Economic Co-operation and Development* (OECD), 2015 erstmals die kollaborative Problemlösekompetenz bei 15-jährigen Schülerinnen und Schülern am Ende ihrer Pflichtschulzeit. Gemäß den übergeordneten Messzielen in PISA betrachtet die OECD somit kollaborative Problemlösekompetenz als Voraussetzung erfolgreicher gesellschaftlicher Teilhabe (OECD, 2017a). Kurzgefasst definiert PISA diese Kompetenz als die Fähigkeit, sich effektiv und zielorientiert am Lösungsprozess einer Problemsituation mit mindestens zwei Akteuren zu beteiligen (OECD, 2017a, S. 134; vgl. Kapitel 1.5). Der vorliegende Beitrag hat das Ziel, die deutschen Ergebnisse zu kollaborativem Problemlösen in der PISA-Studie 2015 im internationalen Kontext sowie weitere, für Deutschland spezifische Analysen vorzustellen. Darüber hinaus ist es ein Ziel dieses Beitrags eine psychometrische Validierung des Testinstruments anhand einer von der internationalen PISA-Erhebung unabhängigen Stichprobe in Deutschland vorzunehmen und die Qualität des Instruments zu diskutieren.

Bei PISA werden Kompetenzen auf individueller Ebene erhoben. Da kollaboratives Problemlösen per se ein interpersonaler Prozess ist, mag das Ansinnen, dieses als individuelle Kompetenz zu messen, paradox oder gar aussichtslos erscheinen. Wie lässt sich der individuelle Beitrag zu einer Gruppenleistung bemessen, die von reziprok wirkenden Bedingungen beeinflusst wird? Hüffmeier und Hertel (2011) zeigten etwa, wie sich die individuelle Leistung von Staffelschwimmern in Abhängigkeit von der Gruppenkonfiguration änderte. Anfang 2017 versuchte das Projekt *sub2*, dem Läufer Eliud Kipchoge durch eine kollaborative Leistung zur Unterbietung der magischen Grenze von zwei Stunden Laufzeit bei einem Marathon zu verhelfen, indem es ihm unter anderem eine Gruppe von Windschatten spendenden Athleten zur Seite stellte. Seitdem versuchen Sportwissenschaftler (etwa Tucker, 2017) zu berechnen, wie Kipchoges Leistung unter Wettkampfbedingungen ausgefallen wäre, das heißt nach Subtraktion des Effekts der kollaborativen Anstrengung. Ähnlich ging die OECD bei der Testkonstruktion vor, da beim Test zum kollaborativen Problemlösen der individuelle Beitrag gemessen werden soll, der in einer gemeinschaftlichen Anstrengung erbracht wird.

Eine aktuell verbreitete Lösung des scheinbaren Paradoxons, die entsprechend in PISA 2015 umgesetzt wurde, ist die simulationsbasierte Erhebung am Computer mit programmierten Teammitgliedern (vgl. Kyllonen, Zhu & von Davier, 2017, S. 6). Dabei können verschiedene Teamkonstellationen (etwa Teamgröße; vgl. Salas, Reyes & Woods, 2017, S. 26), Randbedingungen und Problemstellungen konfiguriert werden, um die

Komplexität möglicher Interaktionen im Testergebnis zu reflektieren (Graesser, Dowell & Clewley, 2017, S. 75) – so können zum Beispiel über die Aufgaben hinweg das Geschlecht der Teammitglieder wie auch die Teamhierarchie variiert werden. Auf diese Weise wird der innerhalb eines vorprogrammierten Szenarios nur additiv wirkende individuelle Beitrag über mehrere Bedingungen aggregiert und somit versucht, den individuellen Beitrag reziproker Prozesse zu approximieren. Die zugrundeliegende Annahme, dass sich solche Wechselwirkungen eindimensional erheben lassen, gilt es empirisch zu prüfen. In der Testung agieren die Schülerinnen und Schüler im Problemraum und interagieren mit simulierten Akteuren zumeist via Chat, indem sie aus vorgegebenen Antworten wählen. Die mit solchen Tests generierten Daten stellen ein Novum für die Bildungsforschung dar und sind mit methodischen Herausforderungen, etwa der stärkeren Interdependenz von Items, verbunden (vgl. von Davier, 2017).

Im Folgenden werden zunächst das von PISA spezifizierte theoretische Rahmenmodell sowie die Methoden zur Messung dieses Konzepts beschrieben. Anschließend stellt der Artikel die internationalen Ergebnisse mit einem Fokus auf Schülerinnen und Schüler in Deutschland vor. Um die nationalen Bedingungen in der Ergebnisdarstellung differenziert zu reflektieren, werden national erhobene Daten hinzugezogen, die über die internationale Datenbasis hinausgehen (z. B. Schulart). Psychometrische Methoden der *Item Response Theory* (IRT) liefern außerdem neue Erkenntnisse zur Messqualität des Instruments. Abschließend werden die Befunde unter anderem im Kontext der methodischen Herausforderungen diskutiert.

## 1 Komplexes Problemlösen im Kontext der Kollaboration

Die Domäne des kollaborativen Problemlösens setzt sich aus der Aufgabenstellung des komplexen Problemlösens einerseits und aus dem kollaborativen Kontext andererseits zusammen. Um ein besseres Verständnis von der in PISA vorgenommenen Operationalisierung zu schaffen, stellt das folgende Kapitel zunächst eine Beispielaufgabe dar. Anschließend skizziert es den aktuellen Stand der Forschung der beiden Komponenten. Der Abschluss des Kapitels zeigt, wie die beiden Bereiche zum kollaborativen Problemlösen vereint werden und welche Limitationen die PISA-Umsetzung dabei, meist aus pragmatischen Gründen, aufweist.

### 1.1 Beispielaufgabe

Zur Veranschaulichung des PISA-Tests zum kollaborativen Problemlösen stellt Abbildung 1 **Fehler! Verweisquelle konnte nicht gefunden werden.** ein prototypisches Item der Unit *Der Besuch* in der computerbasierten Erhebungsumgebung dar (OECD, 2015, S. 59).<sup>1</sup> Das simulierte Szenario handelt von dem Besuch einer ausländischen Schülergruppe. Im ersten Teil der Unit arbeitet die zu testende Schülerin bzw. der zu testende Schüler mit drei simulierten Klassenkameraden zusammen, um ein angemessenes lokales Ausflugsziel zu finden, das für die Gäste interessant sein könnte. Um letztlich eine Empfehlung aussprechen zu können, muss das Team alle Präferenzen diskutieren, ein Missverständnis bezüglich der Öffnungszeiten eines Ausflugsziels aus dem

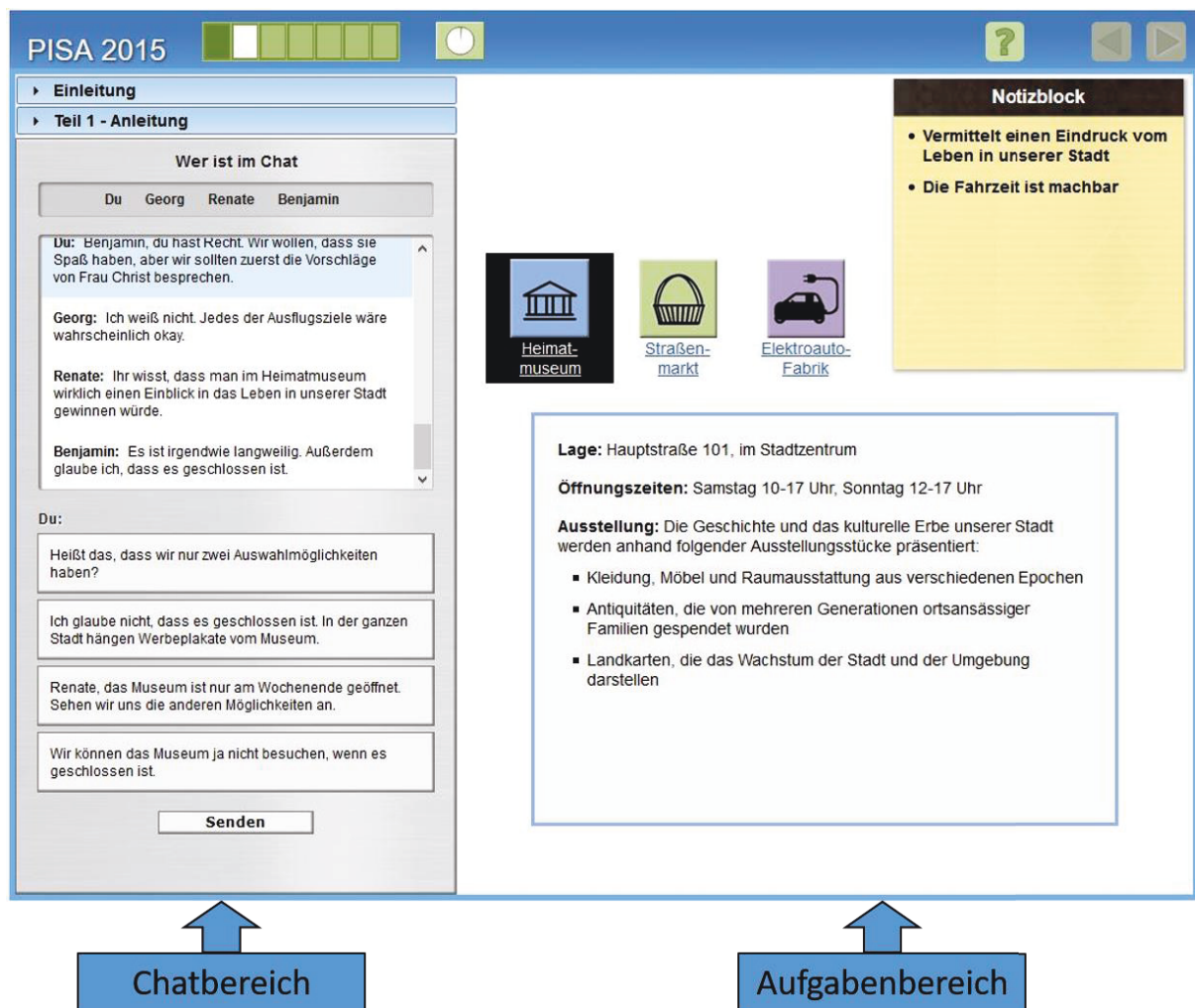
---

<sup>1</sup> Weitere ins Deutsche übersetzte Items finden sich im Internet: <http://www.pisa.tum.de/beispielaufgaben/>

Weg räumen und eine Wahl treffen. Im Beispiel ist zu sehen, wie jedes Teammitglied seine Meinung bezüglich der Ausflugsziele äußert. Georg meint, jedes der Ziele sei akzeptabel, Renate schlägt das Heimatmuseum vor, und Benjamin sagt, dass das Museum langweilig und ohnehin geschlossen sei. Für die Lösung dieses Items muss die Schülerin beziehungsweise der Schüler erstens auf den Weblink des Heimatmuseums klicken, um dessen Öffnungszeiten prüfen zu können, und zweitens die entsprechende Antwort an das Team auswählen. Für die volle Punktzahl müssen sie auf Renates Vorschlag eingehen und aufklären, dass das Museum an dem avisierten Tag geschlossen ist. Tabelle 1 stellt die Antwortkodierung des Items dar (OECD, 2015, S. 59). Eine solche Selektion einer Chatnachricht stellt den häufigsten Itemtypen im Instrument dar. Ein weiteres Beispiel ist die Diskurseröffnung im obengenannten Szenario, bei der Georg fragt, „Okay, womit fangen wir an?“ und die getestete Person wiederum zwischen vier Nachrichten auswählen soll: (1) „Fragen wir Frau Christ, was wir tun sollen.“, (2) „Wir haben drei Möglichkeiten. Stimmen wir ab.“, (3) „Vielleicht sollten wir zuerst eine Weile darüber nachdenken und später darüber diskutieren?“, (4) „Besprechen wir, was für einen guten Ausflug notwendig ist.“ (Zentrum für Internationale Vergleichsstudien, 2017, S. 4). In dem Szenario *Xandar* ist es aber neben dem Voranbringen der Chat-Diskussion beispielsweise auch einmal die Aufgabe, Fragen in einem Schultest zu beantworten, indem die PISA-Schülerinnen und Schüler, Informationen in einer Abbildung durch Anklicken aufrufen (Zentrum für Internationale Vergleichsstudien, 2017, S. 55).

**Tabelle 1.** Antwortkodierung des Beispiel-Items aus der Unit *Der Besuch* (OECD, 2015, S. 59)

<b>Item</b>	CC101106
<b>Kodierung der Antwort</b>	<p>Der/die Schüler/in muss auf den Weblink des Museums klicken, siehe Abbildung 1.</p> <p><b>Volle Punktzahl:</b> Renate, das Museum ist nur am Wochenende geöffnet. Sehen wir uns die anderen Möglichkeiten an.</p> <p><b>Teilpunktzahl:</b> Wir können das Museum ja nicht besuchen, wenn es geschlossen ist.</p>
<b>Klassifikation</b>	(C1) sich mit Teammitgliedern über die aktuell ausgeführten und auszuführenden Handlungen absprechen



**Abbildung 1.** Beispielbildschirm mit Chat- und Aufgabenbereich eines Items aus dem PISA-Test zum kollaborativen Problemlösen; Unit *Der Besuch* (OECD, 2015, S. 59)

## 1.2 Komplexes Problemlösen

Die Ausgangsbasis des PISA-Konstrukts *kollaboratives* Problemlösen stellt *komplexes* Problemlösen dar. Von komplexem Problemlösen spricht man, wenn es sich um Problemstellungen im Hinblick auf dynamische Systeme mit hochgradig vernetzten und anfangs intransparenten Variablen handelt (Dörner & Funke, 2017; Greiff & Fischer, 2013). Das Lösen komplexer Probleme unterscheidet sich vom Lösen sogenannter *analytischer* Probleme in einem wesentlichen Punkt (vgl. Leutner, Fleischer, Wirth, Greiff & Funke, 2012): Bei analytischen Problemen können alle für die Lösung relevanten Informationen aus der Problembeschreibung erschlossen werden; bei komplexen Problemen muss dagegen erst ein Großteil der relevanten Informationen anhand explorierender Interaktionen mit der Problemsituation erzeugt werden. Mit anderen Worten: Bevor man eine Lösung für die Problemsituation finden kann, muss man sich Informationen beschaffen, indem man in die Situation eingreift und zum Beispiel einzelne Parameter variiert und den Einfluss der Variation auf andere Parameter beobachtet. Im Hinblick

auf die PISA-Beispielaufgabe (Abb. 1) müssen die Schülerinnen und Schüler zum Beispiel erst herausfinden, wie die Präferenzen im Team verteilt sind und welche Öffnungszeiten das Heimatmuseum hat. Bei anderen Problemstellungen liegen bestimmte relevante Sachinformationen unterschiedlichen Teammitgliedern vor, die erst gesammelt werden müssen.

In der Forschung zum komplexen Problemlösen dominierte lange Zeit eine experimental-psychologische Perspektive, bei der es zum Beispiel darum ging, schwierigkeitsgenerierende Faktoren und grundlegende kognitive und emotionale Prozesse zu untersuchen. Spätestens mit den PISA-Studien etablierte sich aber eine breite differenzialpsychologisch-psychometrische Perspektive, bei der es darum geht, Problemlösen als Kompetenz zu verstehen, diese Kompetenz messbar zu machen und individuelle Unterschiede zu untersuchen. So konnte etwa gezeigt werden, dass die Kompetenz zur Lösung komplexer Probleme – obwohl deutlich mit Intelligenz korreliert (vgl. die Metanalyse von Stadler, Becker, Gödker, Leutner & Greiff, 2015) – dennoch aber inkrementelle Validität bei der Vorhersage schulischer Leistungen besitzt, wenn Intelligenz als schlussfolgerndes Denken operationalisiert wird (Kretzschmar, Neubert, Wüstenberg & Greiff, 2016). Dabei wird von Dörner und Funke (2017) allerdings kritisch hinterfragt, inwieweit die in solchen Studien verwendeten psychometrischen Tests (bestehend aus Items, die nur *minimal* komplexe Systeme repräsentieren; Funke & Greiff, 2017) tatsächlich noch das erfassen, was ursprünglich mit komplexem Problemlösen gemeint war. So zeigt etwa die PISA-Beispielaufgabe, dass die Komplexität des erforderlichen Explorierens und analytischen Schlussfolgerns im Instrument vergleichsweise gering und geradlinig ist (Aufruf einer Webseite, Erkennen der Öffnungszeiten, Identifizieren des Konflikts mit den Zieldaten). Tatsächlich fügt das Instrument aber weitere Komplexität durch den Kollaborationskontext hinzu, was den Problemraum deutlich erweitert.

### **1.3 Der Kontext der Kollaboration**

Der Kontext der Kollaboration kommt dann zustande, wenn mindestens zwei Individuen mit unterschiedlichen Rollen gemeinsam das Ziel haben, ein Problem zu lösen, was einen wechselseitigen Interaktionsprozess erfordert (Salas, Dickinson, Converse & Tannenbaum, 1992, S. 4). Das PISA-Instrument zum kollaborativen Problemlösen evoziert dies durch die dynamische Szenariendarbietung, bei der stets ein Team ein gemeinsames Ziel hat. Grundsätzlich lassen sich in einem Team drei Kernmechanismen und -ressourcen identifizieren, die die Teameffektivität determinieren (Kozlowski & Ilgen, 2006, S. 116): Kognitionen (wahrgenommenes Teamklima, mentale Modelle des Teams, transaktionales Gedächtnis), Motivation und Emotionen (Teamkohäsion, wahrgenommene Teamwirksamkeit und -stärke) sowie Verhalten (Kompetenzen, Funktionen und regulierende Mechanismen im Team). Das konzeptuelle PISA-Framework zum kollaborativen Problemlösen (OECD, 2017a) und dessen Umsetzung beziehen einige dieser Komponenten stärker ein als andere. So sind etwa mehr als die Hälfte der Items dem Schaffen oder Aufrechterhalten eines gemeinsamen Verständnisses zugeordnet (in obiger Klassifikation: Kognitionen, mentale Modelle). Andere Komponenten hingegen, wie Gespräche, die nicht der unmittelbaren Problemlösung, sondern der Teamkohäsion dienen könnten (Motivation und Emotionen), werden im Test als nicht-kollaborativ gewertet.



Im Unterricht findet Kollaboration oft in Form von Arbeit in Kleingruppen statt. Daran sind meist zwei Erwartungen an den Lernerfolg geknüpft. Zum einen wird durch die vertiefte kognitive Verarbeitung bei Gruppenaktivitäten ein positiver Effekt auf den Lernerfolg der einzelnen Gruppenmitglieder erwartet (King, 2007), die beispielsweise beim gegenseitigen Fragenstellen und Erklären, Argumentieren oder aufeinander aufbauenden Interagieren stattfindet (Chi & Wylie, 2014; Webb et al., 2009). Zum anderen wird von einem Aufbau kognitiver Schemata über den Ablauf von kollaborativem Arbeiten bei den Schülerinnen und Schülern ausgegangen, welche sie in neuen Situationen wieder abrufen und adaptieren können (Fischer, Kollar, Stegmann & Wecker, 2013). Viele der im PISA-Test enthaltenen Szenarien führen solche für Kleingruppenarbeiten typischen Kollaborationsinteraktionen herbei.

#### **1.4 Kollaboratives Problemlösen**

Woher stammt nun das Interesse, die Kompetenz des komplexen Problemlösens im Kontext der Kollaboration zu erfassen? In den letzten Jahrzehnten haben sich die beruflichen Aufgabenspektren hin zu komplexen, nicht-routinierten Anforderungen entwickelt (Autor, Levy & Murnane, 2003; Goos, Manning & Salomons, 2009), und Teamarbeit ist stärker in den Fokus gerückt (Andelfinger & Hänisch, 2017).

Das kollaborative Problemlösen erregt daher zurzeit vor allem im internationalen Kontext Aufmerksamkeit, etwa im von der Industrie geförderten Programm *Assessment and Teaching of 21st Century Skills* (ATC21S; Griffin, McGaw & Care, 2012) oder im Bildungsbericht des Weltwirtschaftsforums (World Economic Forum, 2015). Zudem wird auch in der deutschen Bildungspolitik das Arbeiten in Gruppen aktuell im Zusammenhang mit dem Thema der Digitalisierung von Bildung aufgegriffen. So schreibt die Kultusministerkonferenz zu dem Ziel des planvollen Einsatzes digitaler Lernumgebungen im Unterricht, dass diese den Schülerinnen und Schülern helfen könnten, „sich im Team zu organisieren [und] gemeinsam Lösungen zu entwickeln“ (Kultusministerkonferenz = KMK, 2016, S. 13) – eine Sichtweise, die die KMK bereits im Kontext des verwandten didaktischen Konzepts des kooperativen Lernens zum Ausdruck brachte (KMK, 2012, S. 4). Obwohl nicht expliziert, knüpft die in PISA erhobene kollaborative Problemlösekompetenz an das kooperative Lernen an (Sharan, 1994). Bei beiden Konzepten ist eine positive Interdependenz der gemeinsam Arbeitenden charakteristisch (d. h. gemeinsame Ziele und der individuelle Erfolg sind an den Erfolg der Gruppe geknüpft, etc.) wie auch die Evozierung sozialer und kognitiver Prozesse bei interpersonellen Interaktionen mit dem Ziel, ein gemeinsames Problem zu lösen.

Rund um kollaboratives Problemlösen entstehen aktuell vielzählige Arbeiten zu methodischen Fragen (z. B. Liu, Hao, von Davier, Kyllonen & Zapata-Rivera, 2016), die hier nur exemplarisch angerissen werden, um aufzuzeigen, wie interdisziplinär der Forschungsbereich aktuell aufgestellt wird. So werden etwa Prinzipien der Aufgabenkonstruktion untersucht (z. B. Scoular, Care & Hesse, 2017), und häufig prüfen die Studien die Vergleichbarkeit der Kommunikation in Problemlösesituationen zwischen Menschen mit der Kommunikation zwischen Menschen und simulierten Computeragenten (z. B. Rosen, 2017; Rosen & Mosharraf, 2016; Scoular et al., 2017). Weiterhin werden Methoden vorgeschlagen, um etwa mit der Komplexität der resultierenden Daten oder mit

den Abhängigkeiten der Antworten innerhalb eines Szenarios umzugehen. Dabei werden neu aufkommende Paradigmen wie *Computational Psychometrics* ins Spiel gebracht (von Davier, 2017), die unter anderem Psychometrie mit Maschinenlernen kombinieren. Das folgende Unterkapitel beschreibt, wie kollaborative Problemlösekompetenz in der PISA-Studie 2015 auf dieser bunten Ausgangsbasis mit vielfältigen methodischen Ansätzen definiert, konzeptualisiert und operationalisiert wurde.

### 1.5 Konzeption der kollaborativen Problemlösekompetenz bei PISA 2015

Nach Integration verschiedener früherer theoretischer Darstellungen (Griffin et al., 2012; O'Neil, Chuang & Baker, 2010; Salas, Cooke & Rosen, 2008) entwickelte PISAs Expertengruppe ein Rahmenmodell zur Messung kollaborativer Problemlösekompetenz, auf dem die Instrumentenentwicklung basiert wurde (OECD, 2017a). Gemäß diesem PISA-Rahmenmodell ist eine Schülerin oder ein Schüler im kollaborativen Problemlösen kompetent, wenn sie oder er die Fähigkeit hat, sich effektiv und zielorientiert am Lösungsprozess einer Problemsituation mit mindestens zwei Akteuren zu beteiligen, bei dem die Akteure das für die Lösung notwendige gemeinsame Verständnis herstellen und die notwendige gemeinsame Anstrengung aufbringen sowie ihr Wissen, ihre Fertigkeiten und ihre Anstrengungen vereinen (OECD, 2017a, S. 134). Die theoretische Konzipierung der Domäne schließt somit nahtlos an jene des in PISA 2012 erhobenen komplexen Problemlösens an und bettet sie in einen kollaborativen Kontext ein. Beim komplexen Problemlösen bearbeiten die Schülerinnen und Schüler Probleme am Computer, deren Lösung ausschließlich durch dynamische Interaktion mit der Problemumgebung möglich ist (OECD, 2013).

Für das komplexe Problemlösen unterscheidet PISA 2012 vier involvierte kognitive Prozesse (OECD, 2013, S. 129). (A) Während des *Erforschens und Verstehens* versuchen Problemlösende, eine mentale Darstellung der einzelnen Elemente im Problemraum zu bilden. (B) Beim *Darstellen und Formulieren* bilden sie eine Repräsentation des gesamten Problemraums, zunächst mental, in manchen Fällen auch zusätzlich in grafischen Darstellungen wie Tabellen. (C) Bereiten die Problemlösenden auf dieser Basis Maßnahmen zur Lösung des Problems vor und führen diese aus, spricht PISA von *Planen und Ausführen*. (D) Während des gesamten Prozesses sollten Problemlösende den Fortschritt zur Zielerreichung überwachen und Lösungen aus unterschiedlichen Perspektiven reflektieren (*Überwachen und Reflektieren*).

PISA 2015 behält diese theoretische Konzeption des Problemlösens bei und fügt den Kontext der Kollaboration als zweite Dimension hinzu, durch die drei Hauptkompetenzen kollaborativen Problemlösens beschrieben werden (OECD, 2017a, S. 138). (1) *Gemeinsames Verständnis schaffen und aufrechterhalten*: Die Problemlösenden müssen das gemeinsame Verständnis aller Akteure erkennen, deren verschiedene Perspektiven auf das Geschehen identifizieren und eine gemeinsame Sicht auf die Problemzustände und zu unternehmenden Maßnahmen schaffen können. (2) *Angemessen handeln, um das Problem zu lösen*: Die Art und Weise, wie mit einem Problem in einer Kollaboration am effektivsten umgegangen werden kann, hängt von den Randbedingungen ab. Daher können erfolgreiche Problemlösende die adäquaten Maßnahmen identifizieren, entsprechend der vereinbarten Aufgabenteilung agieren, Fehler korrigieren und

den Erfolg der Maßnahmen evaluieren. (3) *Teamorganisation aufbauen und aufrechterhalten*: Die Problemlösenden müssen mit dem Team eine Struktur aufbauen, die sowohl allen Akteuren passende Rollen zuordnet als auch zur Aufgabenstellung passt. Um vereinbarte Strukturen aufrechtzuerhalten, ergreifen kompetente Problemlösende Maßnahmen, etwa indem sie andere Teammitglieder zur Erledigung ihrer Aufgaben bewegen oder den Erfolg oder Mängel der Teamorganisation rückmelden.

In der Domäne kollaboratives Problemlösen kreuzt PISA 2015 die beiden Dimensionen *komplexes Problemlösen* und *Kollaboration* miteinander (vgl. Tabelle 2), sodass zwölf kollaborative Problemlösefertigkeiten resultieren (z. B. [A1] *Perspektiven und Fähigkeiten der Teammitglieder erkennen*; OECD, 2013, S. 11). Bei der Testkonstruktion wurde angestrebt, verschiedene Aufgabentypen zu entwickeln, die diese Fertigkeiten über das gesamte Instrument hinweg beanspruchen. So leitete die Matrix die Aufgabenkonstruktion und sollte damit sicherstellen, dass alle Aspekte des Rahmenmodells in unterschiedlicher Gewichtung im Messinstrument Entsprechung finden (He, von Davier, Greiff, Steinhauer & Borysewicz, 2017, S. 98). Auf Ebene der Dimensionsausprägungen ist die Gewichtung mit geringen Abweichungen zur Vorgabe des Frameworks (OECD, 2017a, S. 147) gelungen. Einzelne Fertigkeiten der Matrix werden allerdings nur durch sehr wenige oder sogar gar keine Items abgebildet (u. a. A2: 2 Items, A3: 0 Items und D2: 3 Items) und demnach bei Weitem nicht zufriedenstellend abgedeckt (s. Tabelle 2). Somit klafft in dieser Hinsicht eine markante Lücke zwischen dem theoretischen Rahmenkonzept und der Umsetzung. Die einzelnen Dimensionsausprägungen sind sowohl konzeptuell als auch in der Umsetzung im Instrument so stark verzahnt, dass eine separate Auswertung in Form von Subskalen nicht vorgenommen wird. Vielmehr berichtet PISA eine Gesamtskala zur kollaborativen Problemlösekompetenz.

Mit dem Test zum kollaborativen Problemlösen beschreitet PISA Neuland, so dass neben den PISA-typischen Monitoring-Ergebnissen auch die Domäne selbst bzw. deren Operationalisierung Gegenstand von Analysen sein muss. Es stellen sich insbesondere die Fragen, ob die Items eine eindimensionale Skala bilden, ob diese in einer unabhängigen Stichprobe ebenfalls fair misst, wie ihre psychometrische Qualität zu bewerten ist und wie die Skala mit den regelmäßig erhobenen Kompetenzskalen Lesen, Mathematik und Naturwissenschaften zusammenhängt. Während Analysen mit den PISA-Daten, die zu Monitoring-Zwecken erhoben werden, nicht zur Konstruktvalidierung geeignet sind, können sie dennoch die neue Domäne gegenüber den zentralen PISA-Domänen abstecken. Darüber hinaus kann mithilfe einer in Deutschland erhobenen Zusatzstichprobe von Neuntklässlern eine Kreuzvalidierung vorgenommen werden.

**Tabelle 2.** Matrix der kollaborativen Problemlösekompetenzen (OECD, 2017a, S. 137); Anzahl der Items in eckigen Klammern, die in die finale Skalierung eingingen

	(1) Gemeinsames Verständnis schaffen und aufrechterhalten [61 Items]	(2) Angemessen handeln, um das Problem zu lösen [26 Items]	(3) Teamorganisation aufbauen und aufrechterhalten [30 Items]
(A) Erforschen und Verstehen [22 Items]	(A1) Perspektiven und Fähigkeiten der Teammitglieder erkennen [20 Items]	(A2) Art der für die Problemlösung notwendigen kollaborativen Interaktion im Zusammenhang mit den Zielen erkennen [2 Items]	(A3) für die Problemlösung notwendige Rollen verstehen [0 Items]
(B) Darstellen und Formulieren [37 Items]	(B1) Gemeinsame Darstellung des Problems formulieren und dessen Bedeutung im Team aushandeln [24 Items]	(B2) Aufgaben identifizieren und beschreiben [5 Items]	(B3) Rollen und Teamorganisation beschreiben (Kommunikationsprotokoll, Aufgabenaufteilung) [8 Items]
(C) Planen und Ausführen [35 Items]	(C1) Sich mit Teammitgliedern über die aktuell ausgeführten und auszuführenden Handlungen absprechen [5 Items]	(C2) Pläne ausführen [16 Items]	(C3) Aufgabenaufteilung einhalten und aufrechterhalten (z. B. andere Teammitglieder auffordern, ihre Aufgaben zu erledigen) [14 Items]
(D) Kontrollieren und Reflektieren [23 Items]	(D1) Das gemeinsame Verständnis kontrollieren und korrigieren [12 Items]	(D2) Ergebnisse der Handlungen kontrollieren und Erfolg der Problemlösung beurteilen [3 Items]	(D3) Teamorganisation sowie Rollen kontrollieren und anpassen, Teammitgliedern Rückmeldung geben [8 Items]

Da der computerbasierte Test eine im Vergleich zu traditionellen Aufgaben komplexe Simulationsumgebung darstellt, liegt die Vermutung nahe, dass Kompetenzen in Informations- und Kommunikationstechnologien (IKT) eine Rolle für die Lösung der Aufgaben spielen könnten. Daher wird in der vorliegenden Studie außerdem der Zusammenhang des neuen Maßes mit der IKT-Nutzung und -Kompetenz berichtet.

## 1.6 Fragestellungen

Der vorliegende Beitrag adressiert anhand folgender Fragestellungen einerseits die klassischen, deskriptiven Monitoring-Ergebnisse im kollaborativen Problemlösen für Deutschland, um die Bildungspolitik und -forschung etwa über Handlungsbedarf im

High- oder Low-Formerbereich zu informieren, andererseits adressiert der Beitrag die Messgüte der neuen Skala:

**Monitoring.** Wie stellen sich die Ergebnisse von Schülerinnen und Schülern in Deutschland im kollaborativen Problemlösen in Hinblick auf den internationalen Vergleich nach Geschlecht und Kompetenzstufenverteilung sowie im Vergleich zwischen Schularten innerhalb Deutschlands dar?

**Psychometrische Validierung.** Wie hoch fallen die Zusammenhänge der kollaborativen Problemlösekompetenz mit den anderen PISA-Domänen sowie der selbsteingeschätzten Nutzung und Kompetenz von IKT aus? Lässt sich die Qualität des Instruments mittels Kreuzvalidierung anhand einer unabhängigen Stichprobe von Neuntklässlern in Deutschland bestätigen, und wie ist die Messqualität insgesamt zu beurteilen?

## 2 Methode

### 2.1 Stichprobe und Durchführung

PISA 2015 erhob vom 1. März bis 30. Juni 2015 repräsentative Daten von  $n = 535.791$  Fünfzehnjährigen in 72 Staaten, 6.504 davon in Deutschland (für Details s. Sälzer & Reiss, 2016). In Deutschland wurde die Erhebung vom Zentrum für Internationale Vergleichsstudien (ZIB) an der Technischen Universität München koordiniert. Während PISA primär die Kompetenzen in Lesen, Mathematik und Naturwissenschaften testet, wurden die Aufgaben zum kollaborativen Problemlösen in 30 der insgesamt 60 Testformen mit teils verschiedenen, teils überlappenden Aufgaben eingepasst. Neben dem Test wurden die Jugendlichen in einem Fragebogen um Selbstauskünfte zu verschiedenen Kontextbedingungen gebeten. Sowohl der Test zum kollaborativen Problemlösen als auch der zugehörige Fragebogen wurden computerbasiert administriert. In Deutschland bearbeiteten  $n = 1.911$  PISA-Schülerinnen und -Schüler Aufgaben der neuen Domäne. Zusätzlich erhob das deutsche PISA-Zentrum Daten von Neuntklässlern, von denen  $n = 909$  den neuen Test erhielten. Diese unabhängige Zusatzstichprobe setzt sich aus der Ziehung von je 15 Schülerinnen und Schülern einer weiteren neunten Klasse an den PISA-Schulen zusammen. Je ein Datensatz der beiden Stichproben enthielt ausschließlich fehlende Werte in der Testung. International nahmen insgesamt  $n = 124.994$  Schülerinnen und Schüler aus 51 Staaten<sup>2</sup> an der Testung der kollaborativen Problemlösekompetenz teil; 32 dieser Staaten gehören der OECD an.

Aufgrund vielfältiger und über die Bundesländer hinweg uneinheitlicher Strukturen kann bei Vergleichen der Schulart lediglich zwischen der bundesweit einheitlichen Schulart Gymnasium und nicht-gymnasialen Schularten unterschieden werden (vgl. Sälzer & Reiss, 2016, S. 32). Letztere beinhalten Hauptschulen, Schulen mit mehreren Bildungsgängen, integrierte Gesamtschulen und Realschulen. Berufs- und Förderschulen werden nicht separat ausgewiesen, da deren Stichprobenumfänge keine Repräsentativität gewährleisten.

---

<sup>2</sup> Die Populationsabdeckung für Malaysia war nicht ausreichend, sodass die Vergleichbarkeit nicht gewährleistet ist (vgl. OECD, 2017b).

## 2.2 Test zur kollaborativen Problemlösekompetenz

Der Test zum kollaborativen Problemlösen in PISA (OECD, 2017a) besteht aus mehreren Problemszenarien (sog. *Units*), in die sich jeweils mehrere Aufgaben (sog. *Items*) einbetten. Bei der simulationsbasierten, per Computer administrierten Messung kann sich das Instrument an die Antworten der Schülerinnen und Schüler anpassen, indem der Diskurs im Szenario durch verschiedene Antworten unterschiedlich verlaufen kann.

Der Test beinhaltet sechs Units, die sich aus 117 Items zusammensetzen, die größtenteils dichotom (83 %), teilweise aber auch mit Teilpunkten bewertet werden (17 %). Items, die sich als empirisch abhängig voneinander zeigten, wurden zu polytom auswerteten Komposit-Items aggregiert. Dies betraf 14 Einzelitems, die in sechs Komposit-Items zusammengefasst wurden. Items waren in das Szenario einer Unit eingebettet; das heißt, es wurde entweder bewertet, ob die Schülerinnen und Schüler in einer bestimmten Art und Weise im Szenario interagierten (z. B. im Aufgabenbereich eine relevante Webseite öffneten) oder in bestimmter Weise mit den simulierten Gruppenmitgliedern interagierten (d. h., im Chatbereich eine bestimmte Antwort in der Konversation mit den anderen Teammitgliedern auswählten; siehe Abbildung 1). In den meisten Fällen mussten die Schülerinnen und Schüler mit dem Szenario interagieren, damit sich dieses fortsetzte; somit war zumeist das Auslassen eines Items unmöglich (*forced choice*). Im Schnitt benötigten die Schülerinnen und Schüler in Deutschland knapp 12 Minuten, um eine Unit zu bearbeiten, wobei der Median der Bearbeitungszeiten über die Units hinweg von 5 bis 20 Minuten variierte. Auf die Details zur Aufgabenkonstruktion kann hier nicht eingegangen werden (für Details s. Graesser et al., 2017, S. 75). Die Konstruktion musste aber etwa berücksichtigen, dass Distraktoren nicht mit oberflächlichen Merkmalen wie Unhöflichkeit konfundiert waren. Mit den über 50 Teilnehmerstaaten und ihren vielen Testsprachen war es nicht möglich, ein offenes Antwortformat in Kombination mit einer automatischen Auswertung zu verwenden, weshalb zumeist eine Antwort aus drei bis fünf Möglichkeiten ausgewählt werden musste. Die Teamgröße variierte von einem bis fünf simulierten Gruppenmitgliedern.

## 2.3 Die Skala und ihre Kompetenzstufen

Die kollaborative Problemlösekompetenz wird in PISA analog zu den Hauptdomänen auf einer Skala berichtet, deren Mittelwert über alle OECD-Staaten auf 500 festgelegt wird (OECD, 2017b). Genauso wird die Standardabweichung über die OECD-Staaten auf 100 fixiert.<sup>3</sup> Die Skala ist in vier Kompetenzstufen eingeteilt, um zu beschreiben, welche Anforderungen die Schülerinnen und Schüler in einem bestimmten Leistungswertebereich typischerweise bewältigen können (vgl. Tabelle 3). Wie üblich in PISA 2015 (s. dazu Heine et al., 2016), wurde auch die kollaborative Problemlösekompetenz IRT-skaliert und ausgewertet (IRT = *Item Response Theory*; vgl. etwa Rost, 2004). Die Kompetenzstufen sind derart spezifiziert, dass von Schülerinnen und Schülern am unteren Ende einer Stufe erwartet wird, mindestens 50 Prozent eines Aufgabenblocks

---

<sup>3</sup> Die Standardabweichung der Skala kombiniert die Streuungen innerhalb und zwischen den OECD-Staaten. Daher ist eine Standardabweichung innerhalb der Staaten unter 100 Punkten erwartungsgemäß.

zu lösen, der aus gleichmäßig über dieselbe Stufe verteilten Aufgaben besteht (OECD, 2017c, S. 281).

**Tabelle 3.** Überblick zu den typischen Anforderungen der vier Stufen der kollaborativen Problemlösekompetenz (OECD, 2017a, S. 150)

Kompetenzstufe	Wozu die Schülerinnen und Schüler auf der jeweiligen Kompetenzstufe im Allgemeinen in der Lage sind
<b>Stufe 4</b>	Jugendliche auf <b>Stufe 4</b> können komplizierte Problemlöseaufgaben mit hoher Komplexität hinsichtlich der Kollaboration bewerkstelligen. Sie sind in der Lage, Probleme zu lösen, die sich in komplexen Problemräumen mit zahlreichen Randbedingungen befinden, während sie relevante Hintergrundinformationen im Auge behalten. Diese Jugendlichen halten ein Bewusstsein für die Gruppendynamik aufrecht und ergreifen Maßnahmen, um sicherzustellen, dass Teammitglieder gemäß ihrer vereinbarten Rolle handeln. Gleichzeitig sind sie in der Lage, den Fortschritt zur Problemlösung zu überwachen und Hindernisse oder Fehlzustände zu erkennen, die bewältigt werden müssen. Jugendliche auf Stufe 4 ergreifen Initiative und handeln oder erteilen Aufforderungen, um Hindernisse sowie Meinungsverschiedenheiten und Konflikte zu lösen. Sie können die Aspekte der Kollaboration und der Problemlösung bei einer Aufgabe aufeinander abstimmen, effiziente Lösungswege identifizieren und Maßnahmen zur Problemlösung ergreifen.
<b>Stufe 3</b>	Jugendliche auf <b>Stufe 3</b> können Aufgaben bewerkstelligen, die entweder hinsichtlich der Problemlösung oder der Kollaboration komplexe Anforderungen stellen. Diese Jugendlichen können mehrstufige Aufgaben erledigen, die die Integration mehrerer Informationsteile in oft komplexen und dynamischen Problemräumen erfordern. Sie stimmen Rollen innerhalb des Teams ab und identifizieren Informationen, die bestimmte Teammitglieder benötigen, um das Problem zu lösen. Jugendliche auf Stufe 3 können die für eine Problemlösung notwendigen Informationen identifizieren, diese vom entsprechenden Teammitglied einfordern und erkennen, wenn eine bereitgestellte Information falsch ist. Wenn Konflikte auftreten, sind sie in der Lage, den Teammitgliedern bei der Vereinbarung einer Lösung zu helfen.
<b>Stufe 2</b>	Jugendliche auf <b>Stufe 2</b> können zu Kollaborationsbemühungen bei Problemräumen mittlerer Schwierigkeit beitragen. Sie sind in der Lage, bei der Problemlösung zu helfen, indem sie sich mit den Teammitgliedern über die durchzuführenden Handlungen verständigen. Sie sind in der Lage, von sich aus Informationen weiterzugeben, die nicht ausdrücklich von einem anderen Teammitglied angefordert wurden. Jugendliche auf Stufe 2 verstehen, dass nicht alle Teammitglieder über die gleichen Informationen verfügen, und sind in der Lage, bei ihren Interaktionen verschiedene Perspektiven zu berücksichtigen. Sie können dem Team helfen, ein gemeinsames Verständnis dafür zu schaffen, welche Schritte zur Problemlösung erforderlich sind. Die Jugendlichen sind fähig, zusätzliche Informationen für die Problemlösung einzufordern. Sie bemühen sich um die Einigung auf eine Vorgehensweise oder um die Bestätigung dieser durch die Teammitglieder. Jugendliche am oberen Ende von Stufe 2 können die Initiative ergreifen, einen logischen nächsten Schritt oder einen neuen Ansatz zur Lösung des Problems vorzuschlagen.
<b>Stufe 1</b>	Jugendliche auf <b>Stufe 1</b> können Aufgaben mit niedriger Problemkomplexität und begrenzter Komplexität hinsichtlich der Kollaboration bewerkstelligen. Sie können angeforderte Informationen bereitstellen und treffen bei entsprechender Aufforderung Maßnahmen, um Pläne auszuführen. Jugendliche auf Stufe 1 sind in der Lage, Handlungen oder Vorschläge von anderen zu bestätigen. Sie neigen dazu, sich auf ihre eigene Rolle im Team zu konzentrieren. Mit Hilfe der Unterstützung von anderen Teammitgliedern können diese Jugendlichen zur Lösung eines Problems in einfachen Problemräumen beitragen.

## 2.4 Bewertung und Skalierung

In einigen Fällen werden verschiedene Wege in einem Testitem zur kollaborativen Problemlösekompetenz als korrekte Antwort bewertet – diese Wege konvergieren

dann kurze Zeit später im Szenario wieder zum selben Zustand. Anders als bei klassischen Items werden zusätzlich teilweise auch Prozessdaten in die Bewertung der Items mit einbezogen, etwa ob vor der Beantwortung eine gewisse Information abgerufen wurde (was je nach Kontext positiv oder negativ bewertet wird). Aufgrund des *forced-choice*-Formats wurden alle fehlenden Werte tatsächlich als fehlend und nicht als falsch spezifiziert. Die dichotom und polytom bewerteten Items wurden mit dem *2-PL Generalized Partial Credit Model* (GPCM; Muraki, 1992) aus dem Paradigma der IRT mithilfe der Software *mdltm* (von Davier, 2005) eindimensional skaliert, wie dies auch für die anderen Kompetenzdomänen in PISA 2015 der Fall ist (OECD, 2017c). Im Zentrum des 2-PL-Modells stehen die itemspezifischen Schwierigkeits- und Diskriminationsparameter. Erstere geben an, wie schwierig ein Item zu lösen ist, letztere spezifizieren, wie gut das Item fähigere von weniger fähigen Testpersonen unterscheiden kann (ähnlich der Trennschärfe in der Klassischen Testtheorie). Gleichzeitig stellt der Diskriminationsparameter in der IRT eine Gewichtung der Items für die Schätzung der Personenfähigkeit dar: Ist ein Item wenig diskriminierend, trägt es nur wenig zur Schätzung der Kompetenzausprägung bei. Die Kompetenzausprägungen der Schülerinnen und Schüler wurden wie in PISA üblich mit einer latenten Regression modelliert und je Person zehn imputierte, sogenannte *Plausible Values* (PVs) gezogen (vgl. Heine et al., 2016). Die Reliabilität in Form erklärter Varianz durch das Modell beträgt über die Staaten hinweg  $med = ,78$  ( $min = ,70$ ,  $max = ,83$ ; OECD, 2017c, S. 231).

Für die vorliegende Studie wird zur Kreuzvalidierung eine IRT-Reskalierung mit einer weiteren unabhängigen Stichprobe von Neuntklässlern in Deutschland entsprechend dem beschriebenen internationalen Vorgehen vorgenommen. Dazu werden in Modell I die Itemparameter anhand der Daten der deutschen PISA- sowie Neuntklässlerstichprobe gemeinsam als GPCM geschätzt. Modell II, als *Multiple Group IRT Model*, schätzt zusätzlich für die beiden Gruppen jeweils einen eigenen Mittelwert und eine eigene Standardabweichung (Xu & von Davier, 2006), während Modell III, als *Mixture General Diagnostic Model*, zusätzlich die Variation der Itemparameter zwischen den beiden Gruppen zulässt (von Davier, 2007).

### 3 Ergebnisse

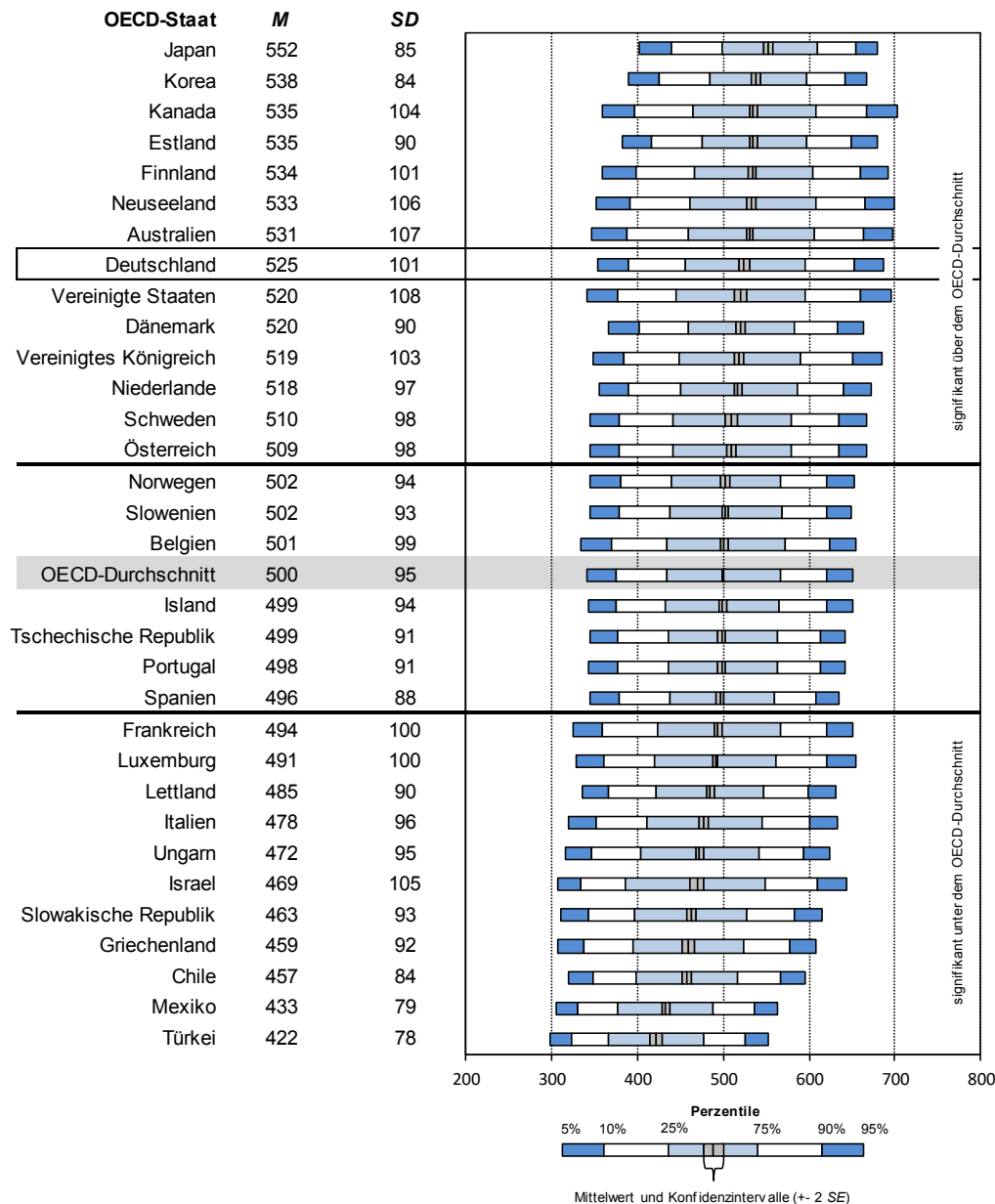
Um die Lesbarkeit zu optimieren, enthält der Ergebnisbericht im Fließtext nicht überall Angaben zu Konfidenzintervallen. Wo im Fließtext angegeben, sind diese in eckigen Klammern als approximierte 95-Prozent-Konfidenzintervalle mit  $\pm 2$  Standardfehlern spezifiziert.<sup>4</sup>

---

<sup>4</sup> Es handelt sich somit also um etwa 95,4-Prozent-Konfidenzintervalle.



Die ersten zwei Unterkapitel stellen die Ergebnisse Deutschlands im Kompetenztest und den Fragebögen in den internationalen Kontext und vergleichen Gymnasien mit anderen Schularten in Deutschland. Die beiden letzten Unterkapitel berichten die Ergebnisse der psychometrischen Validierung inklusive der Zusammenhänge kollaborativer Problemlösekompetenz mit der selbstberichteten Nutzung von und Kompetenz in IKT.



**Abbildung 2.** Perzentilbänder der kollaborativen Problemlösekompetenz in den OECD-Staaten

### 3.1 Kollaborative Problemlösekompetenz: Deutschland im internationalen und Schulart-Vergleich

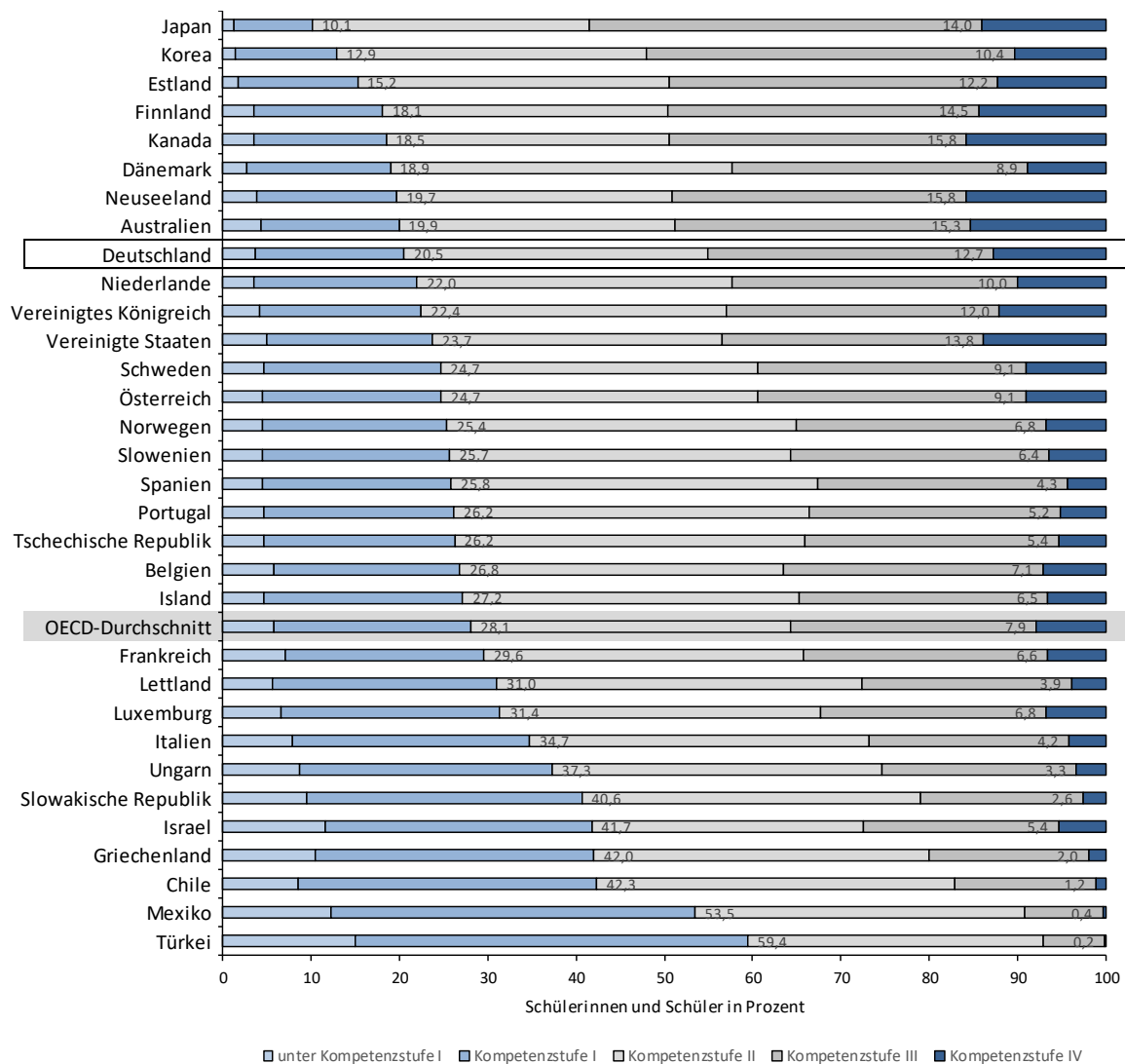
Abbildung 2 zeigt die mittleren Kompetenzausprägungen der untersuchten<sup>5</sup> OECD-Staaten im kollaborativen Problemlösen, deren Streuung sowie Perzentile (für OECD-Partnerstaaten vgl. Tabelle A1 im Anhang). Die Fünfzehnjährigen in Deutschland zeigen mit einem Mittelwert von 525 Punkten [ $\pm 6$ ] eine überdurchschnittliche Kompetenz im Vergleich zur Gesamtheit der untersuchten OECD-Staaten. Den höchsten Kompetenzwert erreicht Japan mit 552 Punkten [ $\pm 5$ ]. Bei den OECD-Staaten befindet sich Deutschland in einer Gruppe mit Australien, den Vereinigten Staaten, Dänemark und dem Vereinigten Königreich, von denen sich der Mittelwert Deutschlands nicht statistisch signifikant unterscheidet. Die Fünfzehnjährigen in Deutschland liegen in der kollaborativen Problemlösekompetenz im Mittel eine viertel Standardabweichung über dem OECD-Durchschnitt und etwa genauso viel unter den leistungstärksten Staaten. Statistisch signifikant unter dem OECD-Durchschnitt liegen unter anderem Luxemburg (491 Punkte) und Griechenland (459 Punkte), während die Türkei mit 422 Punkten [ $\pm 7$ ] das Schlusslicht der OECD-Staaten bildet.

In Deutschland ist der Anteil Fünfzehnjähriger auf der niedrigsten Kompetenzstufe I und darunter sieben Prozent geringer als derselbe Anteil bei den untersuchten OECD-Staaten (vgl. Abbildung 3). Der OECD-Wert von 28 Prozent wird jedoch besonders von dem hohen Anteil niedrig kompetenter Jugendlicher in Mexiko (53 %) und der Türkei (59 %) beeinflusst. Der absolut betrachtet hohe Anteil von 21 Prozent [ $\pm 2$ ] der Fünfzehnjährigen in Deutschland auf Kompetenzstufe I oder darunter ist hingegen nicht vernachlässigbar, beachtet man die niedrigen Anforderungen auf Kompetenzstufe I, die von 17 Prozent [ $\pm 2$ ] erfüllt und von 4 Prozent [ $\pm 1$ ] nicht erfüllt werden. In den meisten OECD-Staaten bewegt sich der Anteil Jugendlicher, die Kompetenzstufe I nicht erreichen, im selben Bereich – lediglich in Japan, Korea und Estland beträgt er weniger als 2 Prozent. Jugendliche auf der höchsten Kompetenzstufe (IV) können üblicherweise schwierige Problemlöseaufgaben mit hoher Komplexität hinsichtlich der Kollaboration bewältigen. Hier liegt der Anteil in Deutschland bei immerhin 13 Prozent [ $\pm 1$ ] und damit fünf Prozent über dem OECD-Wert und vergleichbar mit den Anteilen in den leistungstärksten Staaten (Japan 14 %, Korea 10 %, Kanada 16 %). In durchschnittlich leistungsschwächeren Staaten wie der Türkei und Mexiko erreichen weniger als ein Prozent der Fünfzehnjährigen die höchste Kompetenzstufe.

---

<sup>5</sup> Man beachte, dass die Domäne nicht in allen OECD-Mitgliedsstaaten erhoben wurde. Der besseren Lesbarkeit wegen spricht dieser Beitrag aber zumeist von *den* OECD-Staaten statt von den *untersuchten* OECD-Staaten.

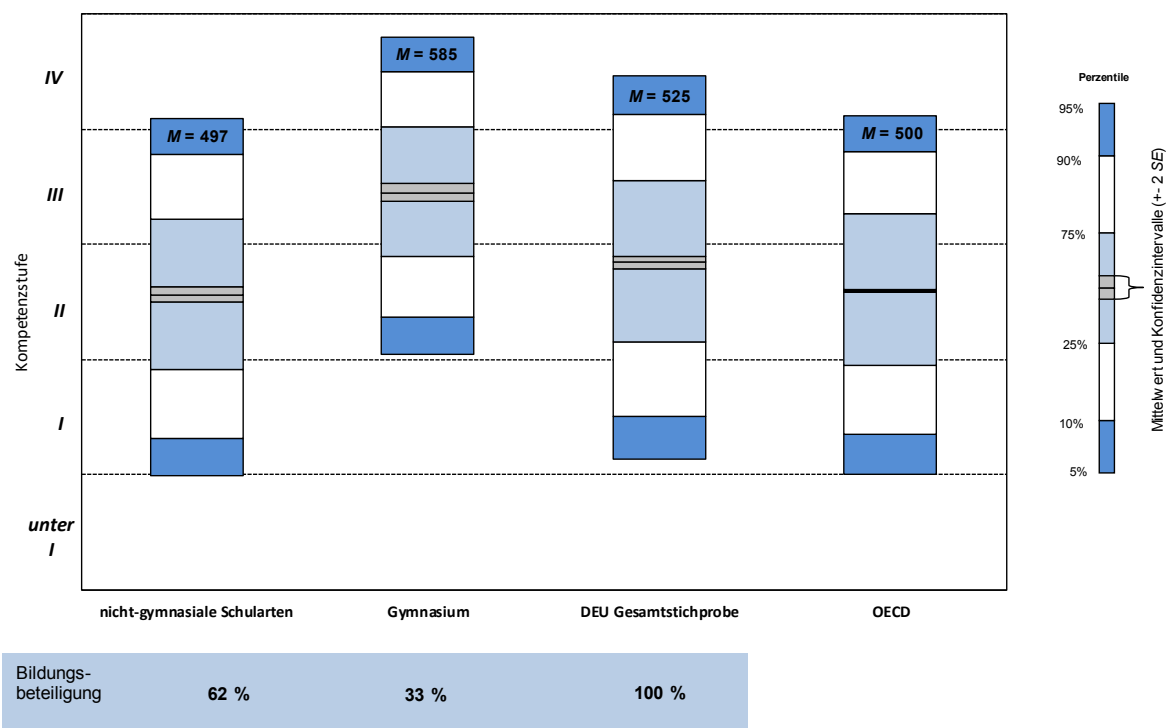
Es zeigt sich ein sehr deutlicher Geschlechtsunterschied zugunsten der Mädchen in der kollaborativen Problemlösekompetenz über alle teilnehmenden Staaten hinweg, welcher bei PISA bisher in ähnlicher Weise nur bei der Lesekompetenz gefunden wurde. Finnland, bekannt für seine Überlegenheit der Mädchen gegenüber den Jungen in allen PISA-Domänen, weist mit 48 Punkten  $[\pm 7]$  den größten Unterschied auf, also knapp eine halbe Standardabweichung – Mexiko und Chile mit je 14 Punkten  $[\pm 6]$  den geringsten. Die Differenz in Deutschland liegt mit 30 Punkten  $[\pm 6]$  im OECD-typischen Bereich. Während Jungen in Deutschland im Mittel 510 Punkte erzielen, liegt das Mittel der Mädchen bei 540 Punkten.



**Abbildung 3.** Prozentuale Anteile von Schülerinnen und Schülern auf Kompetenzstufe I und darunter sowie auf Kompetenzstufe IV

Die Ergebnisse des internationalen Vergleichs machen deutlich, dass unterschiedliche Bedingungen in Personen und im Kontext, wie Geschlecht und Kultur, mit höherer oder niedriger kollaborativer Problemlösekompetenz einhergehen können. Aufgrund der Mehrgliedrigkeit des deutschen Bildungssystems ist in Deutschland auch der Vergleich der Schularten von Interesse.

Schülerinnen und Schüler an Gymnasien erreichen mit 585 Punkten [ $\pm 7$ ] einen höheren Kompetenzwert als an anderen Schularten (497 Punkte [ $\pm 7$ ]). Dabei spiegelt sich die stärkere Heterogenität nicht-gymnasialer Schularten in einer etwas höheren Streuung wider ( $SD = 84$  für Gymnasien und  $SD = 95$  für andere). Besonders informativ ist darüber hinaus die Betrachtung der Verteilung der Schülerinnen und Schüler an nicht-gymnasialen Schularten auf der Kompetenzskala (Abbildung 4). Wie das erste und das vierte Perzentilband zeigen, unterscheidet sich die Verteilung der Schülerinnen und Schüler an nicht-gymnasialen Schulen in Deutschland praktisch nicht von der durchschnittlichen Verteilung der OECD-Staaten. Sowohl die Mittelwerte von 497 und 500 Punkten als auch die Perzentile unterscheiden sich weder praktisch bedeutsam noch statistisch signifikant.



**Abbildung 4.** Perzentilbänder für die kollaborative Problemlösekompetenz für nicht-gymnasiale Schularten, Gymnasium und die Gesamtstichprobe in Deutschland

### 3.2 Psychometrische Validierung: Kreuzvalidierung und Messqualität

Die Ergebnisse der Kreuzvalidierung zur internationalen IRT-Skalierung anhand des Vergleichs der unabhängigen Stichprobe von Neuntklässlern in Deutschland und der deutschen PISA-Stichprobe sind Tabelle 4 zu entnehmen. Wie an den Informationskriterien erkennbar, ist das einfachste Modell I vorzuziehen. Das heißt, dass das Inkludieren der unabhängigen Stichprobe nicht den Bedarf an weiteren Modellparametern nach sich zieht, weder um sparsam die Gruppencharakteristika zu berücksichtigen noch um gruppenspezifische Parameter für jedes Item zu schätzen. Der Zusammenhang zwischen den international verwendeten und den neu geschätzten Itemparametern für die

Neuntklässler-Stichprobe ist stark. Mit  $r = ,91$  bzw.  $r = ,92$  sind die Korrelationskoeffizienten zwischen den Schwierigkeits- bzw. Diskriminationsschätzern der internationalen Skalierung und der Neuntklässler-Kreuzvalidierungsstichprobe vergleichbar hoch zum Zusammenhang zwischen der internationalen Skalierung und der deutschen 15-jährigen-PISA-Stichprobe, die in die internationale Skalierung einging ( $r = ,93$  und  $r = ,93$ ). Absolut betrachtet sind diese Zusammenhänge hoch, für methodische – gegenüber inhaltlich zu interpretierenden – Parametern aber als nicht optimal einzustufen. Ebenfalls enthalten in Tabelle 4 sind die Anzahlen der Items, die jeweils einen auffälligen Modellfit aufweisen. Hier zeigen sich stärkere Modellverletzungen auf Seiten der Neuntklässlerstichprobe in Deutschland als bei der internationalen PISA-Stichprobe, die allerdings noch als moderate Passung einzuordnen sind.

**Tabelle 4.** Ergebnisse der Kreuzvalidierung der IRT-Skalierung

Modell		$n_{\text{Parameter}}$	AIC	BIC	log Penalty	$n$			
						RMSEA>0,05		RMSEA>0,10	
(I)	GPCM	262	183557	185440	0,5898	4		1	
(II)	2 Gruppen mit fixierten Itemparametern	264	183555	185451	0,5898	PISA 5	Neunte 22	PISA 1	Neunte 1
(III)	2 Gruppen mit freien Itemparametern	526	183823	187602	0,5890	PISA 3	Neunte 16	PISA 1	Neunte 1

Anmerkung.  $n_{\text{Parameter}}$ : Anzahl geschätzte Parameter,  $n_{\text{RMSEA}}$ : Anzahl Items ohne guten Modellfit ( $>0,05$ ) bzw. ohne moderaten Modellfit ( $>0,10$ )

Die Ausprägungen der Diskriminationsparameter fallen für das kollaborative Problemlösen markant niedriger aus als für alle anderen Domänen in PISA 2015. Das lässt darauf schließen, dass einzelne Items des neuen Instruments vergleichsweise niedrig mit der Gesamtskala zusammenhängen. 23 Prozent der Items weisen äußerst schlechte Diskrimination von  $\alpha < 0,5$  auf und noch 37 Prozent eine immer noch schwache Diskrimination von  $\alpha < 0,7$ . In der Reskalierung weist sowohl ein Item der deutschen PISA-Stichprobe als auch der Neuntklässler-Stichprobe eine Diskrimination von  $\alpha = 0,0$  auf; das Item wird demnach für die Schätzung der Personenfähigkeit ignoriert. Auf Seiten der Schätzung der Personenfähigkeit bestätigt sich in der Reskalierung jedoch auf Gesamtskalenebene die hohe Präzision der Messung mit einer EAP-Reliabilität von  $Rel = ,90$ .

### 3.3 Die Domäne abstecken: Zusammenhänge mit anderen Kompetenzdomänen und IKT

Da PISA mit dem Test zum kollaborativen Problemlösen eine Domäne mit gänzlich neuen Inhalten und Methoden einbezieht, stellt sich die Frage, wie sich diese im Vergleich mit den bewährten Domänen verhält. Die Lesekompetenz korreliert in PISA 2015 in der OECD zu  $r = ,80$  mit der mathematischen und zu  $r = ,87$  mit der naturwissenschaftlichen Kompetenz. Die letzteren beiden wiederum korrelieren zu  $r = ,88$  untereinander. Diese hohen Zusammenhänge werden seit dem Beginn der PISA-Studien

im Jahr 2000 beständig gefunden. Die kollaborative Problemlösekompetenz reiht sich nun ebenfalls hoch korrelierend ein, jedoch mit etwas geringeren Ausprägungen. Die kollaborative Problemlösekompetenz korreliert mit naturwissenschaftlicher Kompetenz um  $r = ,77$ , mit der Lesekompetenz um  $r = ,74$  und mit der mathematischen Kompetenz um  $r = ,70$ . Für Deutschland fallen die berichteten Korrelationskoeffizienten ohne praktisch bedeutsame Abweichungen ähnlich aus. Im Vergleich dazu korrelierte komplexes Problemlösen in PISA 2012 mit den anderen Domänen gleichermaßen mit Ausnahme der mathematischen Kompetenz, die mit  $r = ,81$  höher lag.

Zu den Zusammenhängen kollaborativer Problemlösekompetenz mit IKT-Selbstberichten ist vorab festzustellen, dass die IKT-Nutzung von den Schülerinnen und Schülern in Deutschland mit knapp einer halben Standardabweichung unter dem OECD-Schnitt als deutlich unterdurchschnittlich berichtet wird; von den OECD-Staaten weisen nur Korea und Japan noch niedrigere Werte auf. Da sich die Zusammenhänge als staatenpezifisch herausstellen, ist über die OECD aggregiert kein praktisch bedeutsamer Zusammenhang der kollaborativen Problemlösekompetenz mit einer der IKT-Kontextvariablen zu finden (schulischer IKT-Gebrauch zuhause:  $r = -,08$ ; IKT-Gebrauch in der Schule:  $r = -,12$ ; selbst eingeschätzte IKT-Kompetenz:  $r = ,05$ ). Während in Deutschland die Zusammenhänge zwischen kollaborativer Problemlösekompetenz mit schulischer IKT-Nutzung zuhause und IKT-Nutzung in der Schule mit jeweils  $r = -,08$  ebenfalls gering ausfallen, gibt es andere Teilnehmerstaaten, in denen diese Variablen etwas stärker zusammenhängen; so korrelieren in Griechenland etwa kollaborative Problemlösekompetenz und IKT-Nutzung in der Schule zu  $r = -,26$ . Bemerkenswert ist außerdem der zumeist negative Zusammenhang mit der IKT-Nutzung in der Schule. Schülerinnen und Schüler, die berichten, häufiger IKT in der Schule zu nutzen, weisen demnach tendenziell schwächere Kompetenzwerte im kollaborativen Problemlösen auf. Das Ausmaß dieser Zusammenhänge ist allerdings gering. Für die schulische Nutzung von IKT zuhause zeichnet sich international ein vielfältigeres Bild. So hängt diese in Chinesisch Taipeh mit  $r = ,15$  leicht positiv mit kollaborativer Problemlösekompetenz zusammen, in einigen anderen Staaten jedoch negativ, etwa in Griechenland ( $r = -,22$ ). Bei der selbstberichteten IKT-Kompetenz findet sich in einigen Staaten wie Deutschland ( $r = -,02$ ) oder Japan ( $r = ,05$ ) kein Zusammenhang mit der kollaborativen Problemlösekompetenz, in anderen Staaten, vor allem mit im Mittel eher unterdurchschnittlichen Kompetenzen bei den 15-Jährigen, zeigt sich hingegen ein leicht positiver Zusammenhang, zum Beispiel in Mexiko ( $r = ,20$ ).

#### 4 Diskussion

Der vorliegende Beitrag ergänzt die internationale Berichtlegung der OECD zur kollaborativen Problemlösekompetenz in PISA um den Fokus auf Deutschland. PISA-Daten bieten Informationen zum Bildungsmonitoring, es werden also Beobachtungen vieler Bedingungen in einer großen, repräsentativen Stichprobe durchgeführt, um einen essentiellen Beitrag zur Informierung der Bildungspolitik, -forschung und -praxis zu leisten. Selbstverständlich verbieten sich Schlussfolgerungen kausaler Art auf Basis der erhobenen Daten. Diese sind genauso wenig legitim wie die Betrachtung von Rankings

hilfreich sind, die den Fokus auf Rangplätze legen anstatt auf Verteilungsmerkmale mit Einbezug von statistischer Genauigkeit und praktischer Bedeutsamkeit.

Die vorliegende Studie berichtet ergänzend, über die Monitoring-Ergebnisse hinausgehend, die Ergebnisse einer psychometrischen Validierung zur Messqualität des Instruments anhand einer Kreuzvalidierung. Anhand der Ergebnisse der psychometrischen Validierung wird zuerst die Messqualität des Instruments diskutiert und beurteilt.

#### 4.1 Qualität des neuen Messinstruments

Der PISA-Test zum kollaborativen Problemlösen mittels simulierter Interaktionen stellt ein Novum dar und muss sich an den geläufigen Gütekriterien empirischer Instrumente messen. Bereits die genaue Betrachtung der Operationalisierung des theoretischen Rahmens überrascht in Hinblick auf die Verteilung der Items (vgl. Tabelle 2) über die zwölf postulierten Fertigkeiten kollaborativen Problemlösens. Offensichtlich wurde bei der Testkonstruktion das Augenmerk primär auf die Verteilung der Items über die verschiedenen kognitiven Aspekte komplexen Problemlösens und die Hauptkompetenzen im kollaborativen Problemlösen („Subskalen“) gelegt. Dass dabei aber eine ganze Zelle völlig unbesetzt bleibt (A<sub>3</sub>: für die Problemlösung notwendige Rollen verstehen), vier Zellen höchstens 5 Items und einer anderen hingegen 24 Items zugeordnet wurden, wirft die Frage auf, ob das Instrument das theoretisch intendierte Konstrukt hinreichend widerspiegelt. Auch ob die Gewichtung der verschiedenen Fertigkeiten theoretisch intendiert ist, scheint fragwürdig. Eine andere Möglichkeit ist, dass dieser Missstand lediglich darauf hinweist, dass die Items nur zwangsweise einzelnen Fertigkeiten zugewiesen wurden, eigentlich aber mehrere Fertigkeiten gleichzeitig beanspruchen. Letzteres liegt nahe, da im Laufe der Szenarien regelmäßig Teamrollen vergeben werden müssen, wofür die Voraussetzung das Verstehen der notwendigen Rollen ist (A<sub>3</sub>).

Die international vorgenommene IRT-Skalierung zeigt bereits auf, dass das Instrument eindimensional im Sinne der *Item Response Theory* misst, wobei für einzelne Item-Staaten-Interaktionen die skalare Messinvarianz gelockert und die Itemparameter freigegeben werden (OECD, 2017c). Die in dieser Studie präsentierten Ergebnisse stützen das internationale Vorgehen insofern, als eine relativ robuste Parameterwiederherstellung für eine unabhängige Stichprobe mit sogar leicht abweichender Populationsdefinition aufgezeigt werden konnte. Nur bei Anlegen des relativ strengen Kriteriums mit  $0,05 \leq RMSEA \leq 0,10$  zeigt sich bei etwa 15 Prozent mehr Items eine Modellverletzung für die Kreuzvalidierungs- als für die originäre PISA-Stichprobe. Insgesamt lässt sich daher schlussfolgern, dass die psychometrische Skalierung auf aggregierter Skalenebene für den PISA-Test zur kollaborativen Problemlösekompetenz robust und zuverlässig ist, was auch eine adäquate Reliabilität der Personenfähigkeitsschätzung untermauert. Letztere steht im Widerspruch zur negativen Seite der berichteten psychometrischen Qualitätskontrolle auf Itemebene, da niedrige Diskriminationsparameter generell zu einer niedrigen Reliabilität führen. Die berichtete EAP-Reliabilität wird hingegen durch die hohe Anzahl an Items sowie vor allem durch die, positiv zu bewertende, gleichmäßige Abdeckung der verschiedenen Schwierigkeitsbereiche beeinflusst. Die von der OECD berichteten Reliabilitätskoeffizienten in Form des Anteils erklärter Varianz durch das Modell (OECD, 2017c, S. 231) zeigt durchaus etwas niedrigere Werte als die anderen Domänen.

Die bemerkenswert schlechten Diskriminationsparameter stellen einerseits die Messqualität einzelner Items in Frage, andererseits wirken sie sich auch negativ auf die Ökonomie des Instruments aus, da Itemscores nur gewichtet mit dem Ausmaß ihrer Diskrimination zum Testscore beitragen. In diesem Sinne sind fast 40 Prozent der verwendeten Items als wenig informativ und wenig ökonomisch einzustufen.

#### **4.2 Monitoring-Ergebnisse**

Im Vergleich zum OECD-Schnitt sind die Schülerinnen und Schüler in Deutschland im kollaborativen Problemlösen überdurchschnittlich kompetent, wobei sie gleichermaßen etwa eine viertel Standardabweichung hinter der OECD-Spitze Japan als auch eine viertel Standardabweichung über dem OECD-Mittel liegen. Während in Deutschland der Anteil der Schülerinnen und Schüler auf der höchsten Kompetenzstufe IV vergleichbar mit jenem in den OECD-Staaten mit den besten Leistungen ist, fällt der Anteil auf der niedrigsten Kompetenzstufe I und darunter etwa doppelt so hoch aus wie in Japan. Obwohl demnach die mittlere Kompetenz der Schülerinnen und Schüler in Deutschland im kollaborativen Problemlösen erfreulich hoch ausgeprägt ist, ist der Anteil niedrigkompetenter Schülerinnen und Schüler im internationalen Vergleich, aber vor allem auch absolut betrachtet, als unerwünscht hoch einzustufen. Dies gilt insbesondere, wenn man sich vergegenwärtigt, welche rudimentären Anforderungen kollaborativ Problemlösende auf Kompetenzstufe I zu bewältigen hatten (vgl. Tabelle 3).

Schülerinnen schneiden in allen Teilnehmerstaaten signifikant besser ab als Schüler. Deutschland liegt hier mit einer Differenz von 30 Punkten im OECD-Mittel. Eine solch konsistente Dominanz der Schülerinnen in einer PISA-Domäne war bislang nur von der Lesekompetenz bekannt. In anderen PISA-Domänen variiert die Geschlechtsdominanz zumeist über verschiedene Länder, was für Lesen und kollaboratives Problemlösen nur im Ausmaß, nicht aber in der Richtung der Fall ist. Interessant ist, dass sich die Geschlechterdifferenz in der Lesekompetenz 2015 erstmalig markant reduziert hatte, was auf die Einführung der computerbasierten Administration zurückzuführen sein könnte (vgl. Weis et al., 2016, S. 268). Für das kollaborative Problemlösen lässt sich dieser Vergleich in Ermangelung einer nicht-digitalen Erhebung nicht ziehen, aber die Befunde zur Äquivalenz der zwischenmenschlichen und Mensch-Maschine-Interaktion in der kollaborativen Problemlösekompetenz (Herborn, in Druck; OECD, 2017b; Rosen, 2017) ließen darauf schließen, dass die Differenz in einer nicht-digitalen Erhebung vergleichbar und nicht noch größer wären. Neben einem potentiellen Einfluss des Administrationsmodus stellen auch die durchschnittlich etwas verschieden ausgeprägten Interessen der Geschlechter potenzielle Erklärungskandidaten dar. Hierüber können teilweise die erhobenen – hier nicht berichteten – Fragebogendaten aufklären. Insgesamt ist aber zu konstatieren, dass bei einer Gesamtzahl von sechs Szenarien im kollaborativen Problemlösen eine mögliche Unausgeglichenheit der Themen einen starken Einfluss haben könnte. Dass sich der neue Test mit der Kollaboration, gegenüber dem klassischen Problemlösen, auf eine Facette fokussiert, die möglicherweise bei den Schülerinnen im Durchschnitt höher ausgeprägt sein könnte als bei den Schülern, wird unten im Kontext der Vorgängerdomäne des komplexen Problemlösens diskutiert.



Besonders beachtenswert erscheint der Befund in Bezug auf den Vergleich von Schularten. Die kollaborative Problemlösekompetenz ist – wie in anderen PISA-Domänen auch – an Gymnasien im Mittel erheblich stärker ausgeprägt als an nicht-gymnasialen Schularten. Allerdings gibt es beim kollaborativen Problemlösen vergleichsweise mehr höherkompetente Schülerinnen und Schüler an nicht-gymnasialen Schularten als dies für andere PISA-Domänen in der Studie 2015 der Fall ist (vgl. Reiss, Sälzer, Schiepe-Tiska, Klieme & Köller, 2016), sodass die Verteilung der kollaborativen Problemlösekompetenz dieser Schülerinnen und Schüler sehr ähnlich zu jener der OECD ist. Dies könnte damit zusammenhängen, dass Jugendliche mit einem privilegierten sozioökonomischen Hintergrund an den Gymnasien überrepräsentiert sind (Müller & Ehmke, 2016; Nold, 2010) und dass der sozioökonomische Hintergrund der Schülerinnen und Schüler zwar auch zum kollaborativen Problemlösen einen erheblichen Zusammenhang hat, dafür aber einen vergleichsweise niedrigeren als bei den anderen PISA-Domänen OECD (2017b). Während in Deutschland die gemeinsamen Varianzanteile des sozioökonomischen Hintergrunds (persönlicher Hintergrund und Hintergrund der Schule) und der Standard-Domänen über 30 Prozent liegen (Naturwissenschaften 34 Prozent, Lesen und Mathematik je 32 Prozent), beträgt dieser Wert für die kollaborative Problemlösekompetenz nur 19 Prozent, was allerdings noch immer leicht über dem OECD-Schnitt von 15 Prozent liegt (OECD, 2017b). Diese Ergebnisse deuten darauf hin, dass bei dem Erreichen hoher kollaborativer Problemlösekompetenzen andere oder zusätzliche Faktoren eine Rolle spielen könnten als bei den wiederholt gemessenen Kompetenzdomänen.

Die Ergebnisse zum kollaborativen Problemlösen sollten außerdem im Lichte der Befunde zum Domänen-Vorgänger, dem komplexen Problemlösen in der PISA-Studie 2012, diskutiert werden. Der internationale Vergleich lieferte 2012 ein ähnliches Gesamtbild wie 2015, jedoch mit einzelnen Ausnahmen. Zu diesen Ausnahmen zählt auch Deutschland, wo die Schülerinnen und Schüler 2012 eine mittlere Kompetenz von 509 Punkten im komplexen Problemlösen aufwiesen (OECD, 2014, S. 52) und sich somit nicht so deutlich vom OECD-Schnitt abhoben, wie das für die 2015 getesteten Schülerinnen und Schüler in Deutschland beim kollaborativen Problemlösen der Fall ist (9 vs. 25 Punkte). Die Geschlechtsunterschiede fielen 2012 beim komplexen Problemlösen ebenfalls über die Staaten hinweg unterschiedlich aus, sind aber dem Betrag nach niedriger als 2015. Im Durchschnitt wiesen Schüler 2012 etwas höhere Kompetenzwerte auf als Schülerinnen (in Deutschland sieben Punkte; OECD, 2014, S. 99). Dies deutet darauf hin, dass der 2015 gefundene ausgeprägte Geschlechtsunterschied in der kollaborativen Problemlösekompetenz zugunsten der Mädchen insbesondere auf kollaborative Fähigkeiten zurückzuführen sein könnte. Der Einfluss des sozioökonomischen Hintergrunds fiel bereits 2012 beim komplexen Problemlösen etwas geringer aus als in den anderen Domänen (OECD, 2014, S. 105), allerdings nicht in dem Ausmaß, wie es nun 2015 beim kollaborativen Problemlösen gefunden wurde. Dass sowohl in nicht-gymnasialen Schularten wie auch bei Schülerinnen und Schülern mit schwächerem sozioökonomischen Hintergrund mehr Höherkompetente im kollaborativen Problemlösen in Deutschland vorzufinden sind, als es im Vergleich zu anderen Domänen zu erwarten war, ist ein erfreuliches Ergebnis.

PISA beansprucht, Qualitätsindikatoren von Bildungssystemen zu messen, daher muss im Rahmen der angeführten Monitoring-Befunde abschließend die Frage gestellt werden, wie kollaborative Problemlösekompetenz nun etwa im Unterricht gefördert werden kann. Zur Unterstützung und Förderung von Kollaboration gibt es verschiedene Ansätze, die in der Forschung zum kooperativen Arbeiten und Lernen teilweise schon sehr umfangreich untersucht wurden. Die Befundlage ist zwar nicht in allen Studienergebnissen eindeutig positiv, jedoch lassen sich Maßnahmen identifizieren, die zumindest im Mittel bedeutsam positive Effekte auf das kooperative Lernen haben. Um beispielsweise das lernförderliche Argumentieren anzuregen, können Lernende auf der Grundlage möglichst weit auseinander liegender Sichtweisen und Lernvoraussetzungen einer kollaborativ arbeitenden Kleingruppe zugeordnet werden (Clark, D'Angelo & Menekse, 2009; Dillenbourg & Hong, 2008). Auch werden sogenannte Kollaborationskripte eingesetzt, um den Gruppenmitgliedern für spezifische kollaborative Szenarien Rollen und Aktivitäten zuzuweisen und die Struktur und Abfolge der gemeinsamen Interaktion so vorzugeben, dass sie möglichst zielführend ist (Vogel, Wecker, Kollar & Fischer, 2017).

### **4.3 Limitationen und Ausblick**

Der Mehrwert, die Genauigkeit und die Gültigkeit sowie die interkulturelle Vergleichbarkeit der abgeleiteten Interpretationen muss Gegenstand weiterer, methodisch orientierter Untersuchungen sein, die eigens dafür weitere Daten zu Außenkriterien generieren oder eine (quasi-)experimentelle Validierung vornehmen. Ein essenzieller Teil einer solchen Validierungsuntersuchung müsste etwa auch den Zusammenhang des neuen Instruments mit dem Konstrukt des komplexen Problemlösens offenlegen. Während die OECD in einer separaten Studie lediglich nachweisen konnte, dass die Verwendung simulierter Agenten im Instrument vergleichbar zur menschlichen Interaktion war (OECD, 2017b), wurde versäumt, das komplexe gemeinsam mit dem kollaborativen Problemlösen zu messen. Die Zusammenhänge des kollaborativen Problemlösens mit den anderen PISA-Domänen geben hier insofern einen ersten Anhaltspunkt, als sich die Messung der kollaborativen Problemlösekompetenz in die bekanntermaßen hohen Interkorrelationen einreicht, wenn sie auch etwas geringer in den Interkorrelationen ausfällt. Daraus kann geschlossen werden, dass die neue Domäne wie intendiert einen Aspekt einbezieht, der im Hinblick auf die berichteten differenziellen Ergebnisse über die bisherigen Indikatoren hinausgehende Befunde liefert. Vertiefende Analysen der vorliegenden Prozessdaten werden solche Untersuchungen zusätzlich informieren können. Die Zusammenhänge des neuen Kompetenzmaßes mit den Selbstberichtsmaßen der Nutzung von IKT und der subjektiven Kompetenz in IKT stellen sich als staaten-spezifisch heraus und zeichnen somit ebenfalls ein komplexes Bild, das weiterer Untersuchung bedarf, um feststellen zu können, inwieweit die Kompetenz im Umgang mit Computern die Messung nicht über das im Rahmenmodell intendierten Sinne hinausgehende Ausmaß beeinflusst. Die OECD (2017b) findet hierzu erste mögliche Anhaltspunkte, dass ein bestimmtes geringes Mindestmaß an IKT-Nutzung und an IKT-Kompetenz gegeben sein müsse, um gute Werte in der kollaborativen Problemlösekompetenz erreichen zu können – ab dieser Schwelle würden die IKT-Aspekte dann keine Rolle mehr spielen. Hinzu kommt jedoch, dass die hier erfassten IKT-Maße lediglich

subjektive Selbstberichte zur IKT-Nutzungshäufigkeit in der Schule sowie subjektive Angaben zur eigenen IKT-Kompetenz sind (z. B. „Wenn sich ein Problem mit einem digitalen Gerät ergibt, denke ich, dass ich es lösen kann.“). Um den Einfluss der Kompetenz im Umgang mit IKT zuverlässig untersuchen zu können, müssten daher vielmehr etablierte objektive Maße der IKT-Kompetenz eingesetzt werden. Genauso wie der Einfluss der IKT-Facetten auf die Messung muss in zukünftigen Arbeiten untersucht werden, inwieweit es dem Test gelingt, objektive und generalisierbare Prädiktionen über soziale Interaktionsprozesse zu treffen, wie der Test ferner mit etablierten Selbstberichtsskalen zur Persönlichkeit oder sozialen Kognitionen zusammenhängt und ob sich das gemessene Konstrukt von generalisierenden kognitiven Grundfertigkeiten unterscheidet. Die Erhebung letzterer wurde von der nationalen Studienleitung in der Erhebung 2015 aufgrund der Umstellung auf computerbasierte Erhebung ausgesetzt und erst nach Erfahrungswerten und Kenntnissen über diese in der Erhebung 2018 wieder eingeführt. Außerdem sollte in zukünftigen Studien die Konstruktvalidität des Instruments zusätzlich durch Zusammenhänge zu sozialen selbstbezogenen Kognitionen sowie sozialen Fähigkeiten gemäß verschiedener Beobachtungsquellen (z. B. Selbstkonzept sozialer Befähigung, Empathie, Persönlichkeitstraits, Peer- oder Lehrerurteile, Situational Judgements) – auch unter Kontrolle individueller Problemlösefähigkeit – untersucht werden.

Die Grenzen der pragmatisch orientierten Operationalisierung kollaborativer Problemlösekompetenz sind offensichtlich. So etwa reduziert das notwendigerweise geschlossene Interaktionsformat den in der Realität höchst komplexen Problemraum in unauthentischem Maße. Mit dem computerbasierten Maß entstehen (gering bis minimal) komplexe Systeme, von denen in reinen Validierungsstudien empirisch gezeigt werden muss, dass sie die prognostische Validität aufweisen, die ihnen die OECD zugedenkt, wenn sie definiert, dass kompetentes kollaboratives Problemlösen im Sinne ihrer Messung Voraussetzung einer erfolgreichen gesellschaftlichen Teilhabe ist. Technologische Innovationen aus dem Bereich der natürlichen Sprachverarbeitung, die inzwischen auch sogenannte Trialoge didaktisch sinnvoll einzusetzen weiß (Graesser, Cai, Morgan & Wang, in Druck), sowie synthetische, aber menschlich agierende Avatare in simulierten Umgebungen (z. B. Kulms & Kopp, 2016) werden die Authentizität der Messung von komplexen Kollaborationssituationen in den nächsten Jahren steigern können. Es gibt bereits erste Ansätze, kollaborative Interaktionen in offenen Kommunikationssystemen automatisch zu kodieren (Hao, Chen, Flor, Liu & von Davier, 2017).

#### **4.4 Conclusio**

Zusammenfassend lässt sich feststellen, dass die OECD mit der Implementierung eines schwierig zu messenden Konstrukts bei PISA innovative Arbeit geleistet hat. Die kommenden Jahre müssen und werden weitere Untersuchungen hervorbringen, welche die interkulturelle Vergleichbarkeit, Genauigkeit, umfassende Validität und Fairness des Instruments bzw. der Interpretation des abgeleiteten Maßes detailliert nachzuweisen haben. Die vorliegende Studie liefert Hinweise darauf, dass das Instrument einigen Qualitätskriterien genügt und sich das Instrument vor allem für die Bildungsforschung als wertvoll herausstellen kann. Andererseits enthält das Instrument markante Schwachstellen in der Operationalisierung des theoretischen Konzepts, zumal Kollaboration als

sehr komplexe Situation auf ein hoch standardisiertes Setting reduziert wird. Vor allem zeigen empirische Qualitätsindikatoren auf Itemebene starke und vermehrte Mängel an, sodass sich die Frage stellt, ob das Instrument für einen weltweiten Einsatz bereits zufriedenstellend ausgereift war und die Ableitung bildungspolitischer Maßnahmen aus den resultierenden Daten erlaubt.

## Danksagung

Das Autorenteam bedankt sich bei dem österreichischen PISA-Zentrum am Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens (bifie) in Salzburg, Österreich, für die Kooperation bei der Übersetzung der Fachterminologie der OECD-Veröffentlichungen ins Deutsche. Diese Passagen sind daher intendiert ähnlich zu den österreichischen Veröffentlichungen.

## Literaturverzeichnis

- Andelfinger, V. P. & Hänisch, T. (Hrsg.). (2017). *Industrie 4.0*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-15557-5>
- Autor, D. H., Levy, F. & Murnane, R. J. (2003). The skill content of recent technological change. An empirical exploration. *The Quarterly journal of economics*, 118 (4), 1279–1333.
- Chi, M. T. H. & Wylie, R. (2014). The ICAP framework. Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49 (4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>
- Clark, D. B., D'Angelo, C. M. & Menekse, M. (2009). Initial structuring of online discussions to improve learning and argumentation. Incorporating students' own explanations as seed comments versus an augmented-preset approach to seeding discussions. *Journal of Science Education and Technology*, 18 (4), 321–333. <https://doi.org/10.1007/s10956-009-9159-1>
- Dillenbourg, P. & Hong, F. (2008). The mechanics of CSCL macro scripts. *International Journal of Computer-Supported Collaborative Learning*, 3 (1), 5–23. <https://doi.org/10.1007/s11412-007-9033-1>
- Dörner, D. & Funke, J. (2017). Complex Problem Solving. What it is and what it is not. *Frontiers in Psychology*, 8 (1153), 1–11. <https://doi.org/10.3389/fpsyg.2017.01153>
- Fischer, F., Kollar, I., Stegmann, K. & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist*, 48 (1), 56–66. <https://doi.org/10.1080/00461520.2012.748005>
- Funke, J. & Greiff, S. (2017). Dynamic problem solving. Multiple-item testing based on minimally complex systems. In D. Leutner, J. Fleischer, J. Grünkorn & E. Klieme (Eds.), *Competence assessment in education. Research, models and instruments (Methodology of Educational Measurement and Assessment, pp. 427–443)*. Cham: Springer International Publishing.
- Goos, M., Manning, A. & Salomons, A. (2009). Job polarization in Europe. *The American Economic Review*, 99 (2), 58–63.
- Graesser, A. C., Cai, Z., Morgan, B. & Wang, L. (2017). Assessment with computer agents that engage in conversational dialogues and trialogues with learners. *Computers in Human Behavior*, 76, 607–616. <https://doi.org/10.1016/j.chb.2017.03.041>

- Graesser, A. C., Dowell, N. & Clewley, D. (2017). Assessing collaborative problem solving through conversational agents. In A. A. von Davier, M. Zhu & P. C. Kyllonen (Hrsg.), *Innovative assessment of collaboration (Methodology of Educational Measurement and Assessment, S. 65–80)*. Cham: Springer.
- Greiff, S. & Fischer, A. (2013). Der Nutzen einer komplexen Problemlösekompetenz. Theoretische Überlegungen und empirische Befunde. *Zeitschrift für Pädagogische Psychologie*, 27 (1-2), 27–39.
- Griffin, P., McGaw, B. & Care, E. (Hrsg.). (2012). *Assessment and teaching of 21st century skills*. Dordrecht: Springer.
- Hao, J., Chen, L., Flor, M., Liu, L. & von Davier, A. A. (2017). CPS-Rater. Automated Sequential Annotation for Conversations in Collaborative Problem-Solving Activities (ETS Research Report No. RR-17-58). Princeton, NJ: ETS. <https://doi.org/10.1002/ets2.12184>
- He, Q., von Davier, M., Greiff, S., Steinhauer, E. W. & Borysewicz, P. B. (2017). Collaborative problem solving measures in the Programme for International Student Assessment (PISA). In A. A. von Davier, M. Zhu & P. C. Kyllonen (Hrsg.), *Innovative assessment of collaboration (Methodology of Educational Measurement and Assessment, S. 95–111)*. Cham: Springer.
- Heine, J.-H., Mang, J., Borchert, L., Gomolka, J., Kröhne, U., Goldhammer, F. et al. (2016). Kompetenzmessung 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. 383–430)*. Münster: Waxmann.
- Herborn, K., Stadler, M., Mustafić, M., & Greiff, S. (in Druck). The Assessment of Collaborative Problem Solving in PISA 2015: Can Computer Agents Replace Humans? *Computers in Human Behavior*.
- Hermann, T., Hirschle, S., Kowol, D., Rapp, J., Resch, U. & Rothmann, J. (2017). Auswirkungen von Industrie 4.0 auf das Anforderungsprofil der Arbeitnehmer und die Folgen im Rahmen der Aus- und Weiterbildung. In V. P. Andelfinger & T. Hänisch (Hrsg.), *Industrie 4.0 (S. 239–253)*. Wiesbaden: Springer Fachmedien Wiesbaden. [https://doi.org/10.1007/978-3-658-15557-5\\_15](https://doi.org/10.1007/978-3-658-15557-5_15)
- Hüffmeier, J. & Hertel, G. (2011). When the whole is more than the sum of its parts. Group motivation gains in the wild. *Journal of Experimental Social Psychology*, 47 (2), 455–459. <https://doi.org/10.1016/j.jesp.2010.12.004>
- King, A. (2007). Scripting collaborative learning processes. A cognitive perspective. In F. Fischer, J. M. Haake, I. Kollar & H. Mandl (Eds.), *Scripting Computer-Supported Collaborative Learning. Cognitive, Computational and Educational Perspectives (Computer-Supported Collaborative Learning, vol. 6, pp. 13–37)*. Boston, MA: Springer Science+Business Media LLC.
- KMK. (2012). Medienbildung in der Schule, Beschluss der Kultusministerkonferenz vom 8. März 2012. Zugriff am 13.07.2017. Verfügbar unter [http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2012/2012\\_03\\_08\\_Medienbildung.pdf](http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2012/2012_03_08_Medienbildung.pdf)
- KMK. (2016). Strategie der Kultusministerkonferenz „Bildung in der digitalen Welt“, Beschluss der Kultusministerkonferenz vom 08.12.2016. Zugriff am 13.07.2016. Verfügbar unter [https://www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/2016/Bildung\\_digitale\\_Welt\\_Webversion.pdf](https://www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/2016/Bildung_digitale_Welt_Webversion.pdf)

- Kozlowski, S. W. J. & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest*, 7 (3), 77–124.
- Kretzschmar, A., Neubert, J. C., Wüstenberg, S. & Greiff, S. (2016). Construct validity of complex problem solving. A comprehensive view on different facets of intelligence and school grades. *Intelligence*, 54, 55–69. <https://doi.org/10.1016/j.intell.2015.11.004>
- Kulms, P. & Kopp, S. (2016). The effect of embodiment and competence on trust and cooperation in human-agent interaction. In *Intelligent Virtual Agents* (S. 75–84).
- Kyllonen, P. C., Zhu, M. & von Davier, A. A. (2017). Introduction. Innovative assessment of collaboration. In A. A. von Davier, M. Zhu & P. C. Kyllonen (Hrsg.), *Innovative assessment of collaboration (Methodology of Educational Measurement and Assessment*, S. 1–18). Cham: Springer.
- Leutner, D., Fleischer, J., Wirth, J., Greiff, S. & Funke, J. (2012). Analytische und dynamische Problemlösekompetenz im Lichte internationaler Schulleistungsvergleichsstudien. *Psychologische Rundschau*, 63 (1), 34–42. <https://doi.org/10.1026/0033-3042/a000108>
- Levy, F. & Murnane, R. J. (2013). Dancing with robots. Human skills for computerized work, third way. NEXT. Zugriff am 13.07.2017. Verfügbar unter <http://content.third-way.org/publications/714/Dancing-With-Robots.pdf>
- Liu, L., Hao, J., von Davier, A. A., Kyllonen, P. & Zapata-Rivera, D. (2016). A tough nut to crack. In Y. Rosen, S. Ferrara & M. Mosharraf (Eds.), *Handbook of research on technology tools for real-world skill development* (pp. 344–359). Hershey, PA: IGI Global. <https://doi.org/10.4018/978-1-4666-9441-5.ch013>
- Mayrberger, K. & Bettinger, P. (2014). Entgrenzung akademischen Lernens mit mobilen Endgeräten. Nutzungspraktiken Studierender in ihrer persönlichen Lernumgebung. In R. Kammerl, A. Unger, P. Grell & T. Hug (Hrsg.), *Jahrbuch Medienpädagogik 11* (S. 155–172). Wiesbaden: Springer. [https://doi.org/10.1007/978-3-658-06462-4\\_9](https://doi.org/10.1007/978-3-658-06462-4_9)
- Müller, K. & Ehmke, T. (2016). Soziale Herkunft und Kompetenzerwerb. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (S. 285–316). Münster: Waxmann.
- Muraki, E. (1992). A generalized partial credit model. Application of an EM algorithm. *Applied Psychological Measurement*, 16 (2), 159–176. <https://doi.org/10.1177/014662169201600206>
- Nold, D. (2010). Sozioökonomischer Status von Schülerinnen und Schülern 2008. *Wirtschaft und Statistik* (2), 138–149.
- O’Neil, H. F., Chuang, S.-h. S. & Baker, E. L. (2010). Computer-based feedback for computer-based collaborative problem solving. In N. M. Seel, P. Pirnay-Dummer & D. Ifenthaler (Eds.), *Computer-Based Diagnostics and Systematic Analysis of Knowledge* (1st ed., pp. 261–279). s.l.: Springer-Verlag. [https://doi.org/10.1007/978-1-4419-5662-0\\_14](https://doi.org/10.1007/978-1-4419-5662-0_14)
- OECD. (2013). *PISA 2012 assessment and analytical framework. Mathematics, reading, science, problem solving and financial literacy (Programme for International Student Assessment)*. Paris: OECD Publishing.
- OECD. (2014). *PISA 2012 results. Creative problem solving (volume V)*: OECD Publishing. <https://doi.org/10.1787/9789264208070-en>

- OECD. (2015). PISA 2015 released field trial cognitive items. Verfügbar unter <https://www.oecd.org/pisa/test/PISA2015-Released-FT-Cognitive-Items.pdf>
- OECD. (2017a). PISA 2015 Assessment and Analytical Framework. Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving. Paris: OECD Publishing. <https://doi.org/10.1787/9789264281820-en>
- OECD. (2017b). PISA 2015 Results (Volume V). Collaborative Problem Solving. Paris: OECD Publishing. <https://doi.org/10.1787/9789264285521-en>
- OECD. (2017c). PISA 2015 technical report. Paris: OECD Publishing.
- Reiss, K., Sälzer, C., Schiepe-Tiska, A., Klieme, E. & Köller, O. (Hrsg.). (2016). PISA 2015. Eine Studie zwischen Kontinuität und Innovation. Münster: Waxmann.
- Rosen, Y. (2017). Assessing students in human-to-agent settings to inform collaborative problem-solving learning. *Journal of Educational Measurement*, 54 (1), 36–53. <https://doi.org/10.1111/jedm.12131>
- Rosen, Y. & Mosharraf, M. (2016). Computer agent technologies in collaborative assessments. In Y. Rosen, S. Ferrara & M. Mosharraf (Eds.), *Handbook of research on technology tools for real-world skill development* (pp. 319–343). Hershey, PA: IGI Global. <https://doi.org/10.4018/978-1-4666-9441-5.ch012>
- Rost, J. (2004). *Lehrbuch Testtheorie Testkonstruktion* (Psychologie-Lehrbuch, 2., überarb. und erw. Aufl.). Bern: Huber.
- Salas, E., Cooke, N. J. & Rosen, M. A. (2008). On teams, teamwork, and team performance. Discoveries and developments. *Human Factors*, 50 (3), 540–547. <https://doi.org/10.1518/001872008X288457>
- Salas, E., Reyes, D. L. & Woods, A. L. (2017). The assessment of team performance. Observations and needs. In A. A. von Davier, M. Zhu & P. C. Kyllonen (Hrsg.), *Innovative assessment of collaboration* (Methodology of Educational Measurement and Assessment, S. 21–36). Cham: Springer.
- Sälzer, C. & Reiss, K. (2016). PISA 2015 - die aktuelle Studie. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (S. 13–44). Münster: Waxmann.
- Scoular, C., Care, E. & Hesse, F. W. (2017). Designs for operationalizing collaborative problem solving for automated assessment. *Journal of Educational Measurement*, 54 (1), 12–35. <https://doi.org/10.1111/jedm.12130>
- Sharan, S. (1994). *Handbook of cooperative learning methods* (The Greenwood educators' reference collection). Westport, Conn.: Greenwood Press.
- Stadler, M., Becker, N., Gödker, M., Leutner, D. & Greiff, S. (2015). Complex problem solving and intelligence. A meta-analysis. *Intelligence*, 53, 92–101. <https://doi.org/10.1016/j.intell.2015.09.005>
- Tucker, R. (2017). Eliud Kipchoge 2:00:25. Zugriff am 13.07.2017. Verfügbar unter <http://sportsscientists.com/2017/05/eliud-kipchoge-20025/>
- Vogel, F., Wecker, C., Kollar, I. & Fischer, F. (2017). Socio-cognitive scaffolding with collaboration scripts: A meta-analysis. *Educational Psychology Review*, 29(3), 477–511. <https://doi.org/10.1007/s10648-016-9361-7>
- Von Davier, A. A. (2017). Computational psychometrics in support of collaborative educational assessments. *Journal of Educational Measurement*, 54 (1), 3–11. <https://doi.org/10.1111/jedm.12129>

- Von Davier, M. (2005). A General Diagnostic Model Applied to Language Testing Data (ETS Research Report No. RR-05-16). Princeton, NJ: ETS.
- Von Davier, M. (2007). Mixture Distribution Diagnostic Models (Bd. 2007, ETS Research Report No. RR-07-32). Princeton, NJ: ETS. <https://doi.org/10.1002/j.2333-8504.2007.tb02074.x>
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54 (3), 427–450.
- Webb, N. M., Franke, M. L., De, T., Chan, A. G., Freund, D., Shein, P., & Melkonian, D. K. (2009). “Explain to your partner”: teachers’ instructional practices and students’ dialogue in small groups. *Cambridge Journal of Education*, 39 (1), 49–70.
- Weis, M., Zehner, F., Sälzer, C., Strohmaier, A., Artelt, C. & Pfof, M. (2016). Lesekompetenz in PISA 2015. Ergebnisse, Veränderungen und Perspektiven. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (S. 249–283). Münster: Waxmann.
- World Economic Forum. (2015). New vision for education. Unlocking the potential of technology. Zugriff am 13.07.2017. Verfügbar unter [http://www3.weforum.org/docs/WEFUSA\\_NewVisionforEducation\\_Report2015.pdf](http://www3.weforum.org/docs/WEFUSA_NewVisionforEducation_Report2015.pdf)
- Xu, X. & von Davier, M. (2006). Cognitive Diagnosis for NAEP Proficiency Data (ETS Research Report No. RR-06-08). Princeton, NJ: ETS.
- Zehner, F., Weis, M., Vogel, F., Leutner, D. & Reiss, K. (2017). *Kollaboratives Problemlösen in PISA 2015. Befundhöhepunkte rund um Deutschland*. München: Technische Universität München. <https://doi.org/10.14459/2017md1401788>
- Zentrum für Internationale Vergleichsstudien. (2017). Beispielaufgabe Kollaboratives Problemlösen. Zugriff am 09.05.2018. Verfügbar unter [http://www.pisa.tum.de/fileadmin/woobgi/www/PISA\\_2015\\_FT\\_CPS.pdf](http://www.pisa.tum.de/fileadmin/woobgi/www/PISA_2015_FT_CPS.pdf)



Tabelle A1. Mittelwerte, Streuungen und Perzentile der kollaborativen Problemlösekompetenz

OECD-Staaten	Perzentile									
	M	(SE)	SD	(SE)	5 %	10 %	25 %	75 %	90 %	95 %
Japan	552	(2,7)	85	(1,8)	402	440	499	610	655	680
Korea	538	(2,5)	84	(1,5)	390	425	484	598	641	667
Kanada	535	(2,3)	104	(1,0)	358	397	465	607	667	702
Estland	535	(2,5)	90	(1,3)	382	416	475	598	650	679
Finnland	534	(2,6)	101	(1,5)	359	399	466	605	660	693
Neuseeland	533	(2,4)	106	(1,7)	353	391	460	608	666	700
Australien	531	(1,9)	107	(1,3)	347	388	460	607	664	698
Deutschland	525	(2,8)	101	(1,5)	354	390	456	595	653	686
Vereinigte Staaten	520	(3,6)	108	(1,7)	341	376	445	596	659	696
Dänemark	520	(2,5)	90	(1,2)	367	402	460	583	634	663
Vereinigtes Königreich	519	(2,7)	103	(1,1)	348	384	449	591	651	686
Niederlande	518	(2,4)	97	(1,5)	355	389	450	586	640	672
Schweden	510	(3,4)	98	(1,8)	344	379	441	579	635	667
Österreich	509	(2,6)	98	(1,5)	345	379	441	580	635	667
Norwegen	502	(2,5)	94	(1,6)	345	380	439	568	621	653
Slowenien	502	(1,8)	93	(1,3)	345	378	438	568	620	649
Belgien	501	(2,4)	99	(1,4)	334	369	434	572	625	655
<b>OECD-Durchschnitt</b>	<b>500</b>	<b>(0,5)</b>	<b>95</b>	<b>(0,3)</b>	<b>341</b>	<b>375</b>	<b>435</b>	<b>567</b>	<b>621</b>	<b>652</b>
Island	499	(2,3)	94	(1,9)	343	375	433	566	620	652
Tschech. Republik	499	(2,2)	91	(1,4)	344	377	436	563	614	643
Portugal	498	(2,6)	91	(1,3)	343	377	437	562	613	641
Spanien	496	(2,1)	88	(1,1)	345	379	438	559	607	635
Frankreich	494	(2,4)	100	(1,5)	325	359	424	566	620	651
Luxemburg	491	(1,5)	100	(1,0)	328	361	420	561	621	654
Lettland	485	(2,3)	90	(1,3)	335	367	423	547	599	631
Italien	478	(2,5)	96	(1,6)	319	353	412	545	601	633
Ungarn	472	(2,4)	95	(1,6)	316	347	404	541	594	625
Israel	469	(3,6)	105	(1,8)	307	334	386	548	609	643
Slowakische Republik	463	(2,4)	93	(1,5)	311	343	398	528	583	615
Griechenland	459	(3,6)	92	(1,6)	307	338	394	524	578	609
Chile	457	(2,7)	84	(1,3)	319	348	398	516	567	596
Mexiko	433	(2,5)	79	(1,5)	305	331	378	488	536	564
Türkei	422	(3,4)	78	(1,6)	298	323	367	477	526	553

signifikant über dem  
OECD-Durchschnittnicht signifikant verschieden  
vom OECD-Durchschnittsignifikant unter dem  
OECD-Durchschnitt

OECD- Partnerstaaten										
Singapur	561	(1,2)	97	(1,2)	392	432	499	630	680	709
Hong Kong (China)	541	(2,9)	90	(1,6)	382	420	483	604	652	681
Macao (China)	534	(1,2)	90	(1,1)	377	415	476	596	645	672
Chinesisch Taipeh	527	(2,5)	90	(1,5)	370	407	468	590	639	667
B-S-J-G (China)	496	(4,0)	97	(2,1)	333	368	429	564	620	651
Russland	473	(3,4)	92	(1,4)	324	355	410	537	593	626
Kroatien	473	(2,5)	87	(1,5)	328	359	412	534	585	614
Litauen	467	(2,5)	91	(1,4)	319	349	404	532	584	613
Bulgarien	444	(3,9)	98	(1,7)	290	319	370	515	575	606
Uruguay	443	(2,3)	91	(1,3)	301	328	376	506	564	597
Costa Rica	441	(2,4)	78	(1,3)	316	343	387	494	542	570
Malaysia	440	(3,3)	80	(1,7)	310	337	384	495	543	569
Thailand	436	(3,5)	83	(1,7)	307	332	375	492	547	580
Ver. Arabi. Emirate	435	(2,4)	95	(1,0)	289	317	366	500	563	598
Kolumbien	429	(2,3)	83	(1,4)	300	325	370	486	539	571
Peru	418	(2,5)	83	(1,6)	287	313	358	475	529	561
Montenegro	416	(1,3)	79	(1,2)	291	315	359	471	520	548
Brasilien	412	(2,3)	87	(1,3)	277	304	350	470	529	564
Tunesien	382	(1,9)	59	(1,4)	291	310	341	419	459	485

\*B-S-J-G (China) bezieht sich auf vier Provinzen aus China, die an der PISA-Studie teilgenommen haben:  
Peking, Shanghai, Jiangsu und Guangdong.