JEM JOURNAL of EDUCATIONAL MEASUREMENT NCME national council on measurement in education

Journal of Educational Measurement
Spring 2017, Vol. 54, No. 1, pp. 103–122

Mapping an Experiment-Based Assessment of Collaborative Behavior Onto Collaborative Problem Solving in PISA 2015: A Cluster Analysis Approach for Collaborator Profiles

Katharina Herborn, Maida Mustafić, and Samuel Greiff
University of Luxembourg

Collaborative problem solving (CPS) assessment is a new academic research field with a number of educational implications. In 2015, the Programme for International Student Assessment (PISA) assessed CPS with a computer-simulated human-agent (H-A) approach that claimed to measure 12 individual CPS skills for the first time. After reviewing the approach, we conceptually embedded a computer-based collaborative behavior assessment (COLBAS) into the overarching PISA 2015 CPS approach. COLBAS is an H-A CPS assessment instrument that can be used to measure certain aspects of CPS. In addition, we applied a model-based cluster analysis to the embedded COLBAS aspects. The analysis revealed three types of student collaborator profiles that differed in cognitive performance and motivation: (a) passive low-performing (non-)collaborators, (b) active high-performing collaborators, and (c) compensating collaborators.

Worldwide, humans are increasingly faced with technology-rich and ill-defined tasks that demand flexibility and individual problem-solving skills (Greiff, 2012) in the workplace and in other areas of life. Many of these tasks are solved only through collaboration because solving problems collaboratively enables us to divide labor, incorporate information from many sources, and enhance the solutions' quality (Organisation for Economic Co-operation and Development [OECD], 2013). Collaborative problem solving (CPS)[1] enables individuals to communicate about problem situations, join their perspectives and skills, and solve tasks that are difficult or impossible to achieve individually (OECD, 2013). By contrast, individual problem solving, the competency to individually discover problems within unknown systems and apply knowledge to solve them (Greiff, 2012), does not offer these advantages. CPS is a construct that is composed of cognitive individual problem-solving skills and social collaboration skills (O'Neil, Chuang, & Baker, 2010). More specifically, CPS can be defined as "the capacity of an individual to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills and efforts to reach that solution" (OECD, 2013, p. 6).

As sufficient CPS skills are increasingly expected in the workplace (OECD, 2013), not surprisingly collaboration has evolved into a key 21st century skill and a contributor to university and career readiness (Rosen & Tager, 2013). International education commissions have begun to integrate CPS learning and development into school curricula to adequately prepare students for future careers (OECD, 2013). Specifically, pedagogically effective methods for implementing CPS have been developed and refined in recent years by both education commissions and researchers with a strong focus on improving educational technology in the classroom (von Davier & Halpin, 2013). However, assessment approaches for evaluating students' CPS skills as a means of outcome evaluation from educational systems are in a nascent state (cf. von Davier & Halpin, 2013). In line with this idea, the OECD states "there are no established reliable methods for large-scale assessments of individuals solving problems in a collaborative context and no existing international assessments in wide use" (2013, p. 17). Moreover, there is no established statistical theory for the assessment of CPS to date (von Davier & Halpin, 2013).

Recognizing the growing importance of CPS assessment, in its sixth survey in 2015 the Programme for International Student Assessment (PISA) included the large-scale assessment of CPS with a computer-simulated human-agent (H-A) approach for measuring individual-level CPS skills. In PISA 2015, test-takers' CPS skills were assessed while they collaborated with computer-simulated agents in digital group tasks. By including this H-A approach, PISA broke new ground in the large-scale assessment of CPS in educational settings.

To meet the first aim of the current article, we review the PISA 2015 H-A approach. To meet the second and main aim of this article, we embed COLBAS (i.e., computer-based collaborative behavior assessment; Krkovic, Wüstenberg, & Greiff, 2016), an already existing H-A CPS assessment that covers specific aspects of the PISA 2015 approach, into the broad and overarching PISA 2015 H-A approach. We then apply a model-based latent class cluster analysis (Fraley & Raftery, 2002) to empirically analyze the embedded COLBAS aspects in a preexisting COLBAS data set (481 seventh graders in Germany). Therefore, the resulting clusters are based on a preexisting data set generated by COLBAS, and not by the PISA 2015 assessment. A model-based cluster analysis enabled us to identify particular subgroups of CPS behavior (Fraley & Raftery, 2002), which provided information about specific aspects of the PISA 2015 assessment and helped generate ideas for the statistical analysis of the PISA 2015 CPS results. We will present three classified latent cluster profiles and their implications of potential CPS behavior in PISA 2015: (a) passive low-performing (non-)collaborators, (b) active high-performing collaborators, and (c) compensating collaborators. These suggestions for the analyses could be conducted when PISA 2015 CPS results are published, which will be presumably in late 2017.

CPS Assessment in PISA 2015

Since 2000, the triennial PISA assessment cycles have aimed to "assess to what extent students at the end of compulsory education, can apply their knowledge to real-life situations and be equipped for full participation in society" (OECD, 2015a). Participating students are between the ages of 15 years 3 months and 16 years 2 months at the time of the test (OECD, 2015a). The duration of the PISA test is approximately two hours, and it assesses the domains of reading, mathematics, and science literacy. Because transversal skills that extend across domains have increasingly been acknowledged as important for full participation in society (OECD, 2015a), individual problem solving and financial literacy were integrated into PISA 2012[2] as optional assessments (cf. OECD, 2013). As a further comprehensive development of individual problem-solving assessment in PISA 2012, problem-solving skills were assessed on the group level (i.e., CPS) in PISA 2015. PISA 2015 incorporated the assessment of CPS to internationally evaluate students' CPS potential as a means of providing an outcome evaluation of their respective educational systems and their general career readiness. PISA 2015 treated the CPS skills as conjoint constructs of the PISA 2012 individual problem-solving processes and newly conceptualized collaboration dimensions. The processes that originated from the PISA 2012 individual problem-solving assessment include (A) exploring and understanding, (B) representing and formulating, (C) planning and executing, and (D) monitoring and reflecting. These processes were "crossed" with the collaboration dimensions, and conceptualized 12 distinguishable CPS skills, which are illustrated in the PISA 2015 CPS 12-cell matrix (OECD, 2013) as presented in Table 1.[3]

In the matrix (see Table 1), Process A (i.e., exploring and understanding) assesses a student's cognitive processes in exploring an initial problem state and understanding the problem during an exploration phase (OECD, 2013). Process B (i.e., representing and formulating) measures a student's level of selecting and structuring information about the given problem and integrating existing knowledge, as well as formulating and representing the known information, for example, by creating tables (OECD, 2013). Process C (i.e., planning and executing) assesses the extent to which the students plan how to achieve their desired goal state (e.g., setting subgoals; OECD, 2013). Finally, process D (i.e., monitoring and reflecting) captures the progress of the students and their decisions about their solutions (OECD, 2013).

Further, the newly conceptualized collaboration dimensions include (1) establishing and maintaining a shared understanding, (2) taking appropriate action to solve the problem, and (3) establishing and maintaining team organization. The first PISA 2015 collaboration dimension, (1) establishing and maintaining a shared understanding, describes students' understanding of which information is shared, developed, and maintained in the group, which requires them, for instance, to monitor and evaluate contributions, communicate, share knowledge, and perform troubleshooting in the group (OECD, 2013). The CPS skills, A1–D1 in the matrix, are thus conceptualized by crossing the four individual problem-solving processes (i.e., A–D) with the first collaboration dimension of PISA 2015 (i.e., (1) establishing and maintaining a shared understanding; see Table 1). Further, the second collaboration dimension, (2) taking appropriate action to solve the problem, describes the identification and stepwise performance of activities undertaken to solve the problem (OECD, 2013). Thus, the CPS skills A2–D2 are created from the combination of individual problem-solving processes and the second collaboration dimension. Finally, the third collaboration dimension, (3) establishing and maintaining group organization, describes the ability to organize and structure the group around the problem (OECD, 2013) and, along with the individual problem-solving processes, creates the CPS skills A3–D3.

Table 1

The 12-Cell Matrix (in Italics) Outlining the 12 CPS Skills in the PISA 2015 Assessment Drawn From the OECD Draft Report for CPS in PISA 2015 (OECD 2013)

	(1) Establishing and maintaining shared understanding Questioning	*(2) Taking appropriate action to solve the problem* Asserting	*(3) Establishing and maintaining team organization* Requesting
(A) Exploring and understanding **Knowledge acquisition**	*(A1) Discovering the perspectives and abilities of team members*	*(A2) Discovering the type of collaborative interaction needed to solve the problem, along with goals*	*(A3) Understanding roles to solve the problem*
(B) Representing and formulating **Knowledge acquisition**	*(B1) Building a shared representation and negotiating the meaning of the problem (common ground)*	*(B2) Identifying and describing tasks to be completed*	*(B3) Describe roles and team organization (communication protocol/rules of engagement)*
(C) Planning and executing **Knowledge application**	*(C1) Communicating with team members about the actions to be / being performed*	*(C2) Enacting plans*	*(C3) Following rules of engagement (e.g., prompting other team members to perform their tasks)*
(D) Monitoring and reflecting	*(D1) Monitoring and repairing the shared understanding*	*(D2) Monitoring results of actions and evaluating success in solving the problem*	*(D3) Monitoring, providing feedback, and adapting the team organization and roles*

Note. The rows (A), (B), (C), and (D) show the individual problem-solving processes. The columns (1), (2), and (3) show the collaboration dimensions. The 12 cells (A1–D3) illustrate the 12 CPS skills that conceptualized CPS in PISA 2015. The embedded COLBAS aspects, which are relevant for the later cluster analysis, are in bold. COLBAS aspects cannot be theoretically embedded in row D.

In PISA 2015, these CPS skills were assessed at the individual level by employing the computer-simulated H-A approach to standardize assessment conditions when measuring CPS skills. The H-A approach consisted of digital units that varied in length from 5 to 20 min and were summed into an overall CPS score. Each

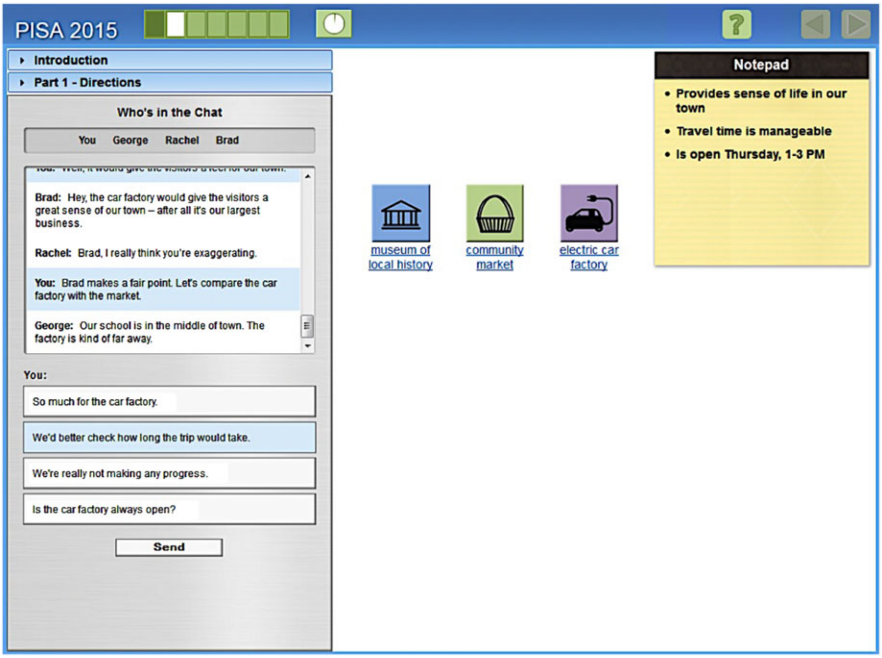

Figure 1. Screenshot of the published PISA 2015 CPS example unit "The Visit" (OECD, 2013, p. 63). The left part of the screen displays the chat window and chat options, and the right part simulates the task space in which actions are performed.

test-taker (TT) completed the units on a computer individually. Addressing the units required efficient collaboration with a minimum of one and a maximum of three computer-simulated agents. The collaboration was based on communication through predefined messages, which were exchanged through a chat box, and actions were implemented in a task space. For example, in the unit "The Visit" (which was incorporated into the PISA 2015 field trial but was not included in the main study), the TT collaborated with the three computer-simulated agents called George, Rachel, and Brad (OECD, 2015b). Figure 1 illustrates the visual representation of the CPS unit "The Visit."

The unit required the TT and agents to choose a local point of interest for a class trip. The TT received four predefined messages that could be used to respond to George, Rachel, and Brad, all being computer-simulated agents. TTs could always view the entire chat history to reread information when needed. TTs were always allowed to choose one of the given messages. Typing messages independently (so-called open chat) was impossible. One predefined message (in rare cases, two) was always scored as the correct choice, reflecting the high-level competency of one or sometimes two particular CPS skills in the matrix. For example, in Figure 1, the response "We'd better check how long the trip will take" is used to score the skill (C1) communicating with team members about the actions to be/being performed. Overall, "The Visit" is used to assess the CPS skills A1–D1, B2–D2, and B3–D3.

Primarily the chat responses but also additional actions including, for instance, clickable links to websites, were scored and summed into an overall CPS score. Selecting predefined messages and implementing actions were mandatory for proceeding through the units (OECD, 2013).

Because the PISA 2015 CPS approach is a wide overarching assessment that covers many aspects of CPS, it allows other existing instruments to be embedded in the measure. One existing computer-assisted assessment for collaborative behavior in students is COLBAS (Krkovic et al., 2016), which also employs the H-A approach and assesses CPS through the exchange of predefined messages. The two approaches overlap conceptually in specific aspects, which allows for the embedding of COLBAS in the PISA 2015 CPS framework. Considering that PISA 2015 results have not been released yet (release is expected to take place in late 2017), the embedding of COLBAS in the PISA 2015 framework enabled us to already measure certain aspects of CPS assessment in PISA 2015 on the basis of COLBAS data. More specifically, we measured profiles of particular collaborative behavior in students as different profiles of individual problem-solving processes and collaboration have been found in human-human (H-H) settings, in which participants collaborate with human collaboration partners instead of agents. For instance, in one study by Chung, O'Neil, and Herl (1999), groups were assigned anonymously, and the group members communicated through predefined messages during a group task (i.e., a knowledge map). The results showed that low-performing groups sent more predefined messages during the knowledge map group task but gained less knowledge than high performing groups, which communicated less but gained more knowledge from their conversations. However, profiles of individual problem-solving processes and collaboration in H-A settings have been more or less disregarded in CPS research, especially in large-scale assessments. The known differences in collaborator profiles researched in H-H settings raise the question of whether collaborator profiles also exist in H-A settings. If so, then PISA 2015 CPS assessment may reveal important insights into potential types of CPS behavior in students.

Computer-Assisted Assessment for Collaborative Behavior (COLBAS)

COLBAS is based on the reliably established computer-based MicroDYN approach (Greiff, Wüstenberg, & Funke, 2012). MicroDYN is a computer-based test for individual problem-solving skills assessesing individual problem solving in two phases. (1) In the knowledge acquisition phase, the TT collects information about the stated problem; and (2) in the knowledge application phase, the TT applies the acquired knowledge to solve the stated problem. The assessment of individual problem solving through knowledge acquisition and knowledge application dimensions is an established method in academic research (for further details, see Greiff et al., 2012). In the knowledge acquisition phase, MicroDYN measures how the TT explores the problem and the input and output variables (e.g., pressing buttons, observing effects) and how well the TT draws connections between them on a graphical model (cf. Wüstenberg, Greiff, & Funke, 2012). TTs achieve full scores for drawing the correct model. In the knowledge application phase, MicroDYN measures how the TT uses the model to achieve a particular goal. TTs are given full scores for achieving the specific goals. More detailed information on MicroDYN can be found in Greiff et al. (2012).

Because COLBAS is based on MicroDYN and extended to collaborative settings, COLBAS units also measure CPS through knowledge acquisition and knowledge application phases. However, to apply MicroDYN to collaborative settings, the knowledge acquisition phase is extended to include collaboration; that is, the information needed to successfully solve the tasks can be fully acquired only by collaborating with a computer-simulated agent. Three collaboration buttons (i.e., questioning, requesting, and asserting) offer lists of several predefined messages via drop-down chat menus and an open chat window for assertions. These buttons enable options for communicating with the agent to exchange information. The TT can ask questions, make assertions, or request necessary information from the agent to help solve the unit. Therefore, COLBAS is conceptually a five-dimensional CPS assessment instrument that measures CPS through two individual problem-solving processes (knowledge acquisition and knowledge application) and three collaboration dimensions (questioning, requesting, and asserting). Figure 2 provides a visual representation of a COLBAS knowledge acquisition phase. As shown in Figure 2, the TT can explore the problem in a task space on the left side of the screen. The right side of the screen contains a collaboration space in which the TT can collaborate with the computer agent by pressing the question, request, and assert bars to select the type of communication with the agent. In each unit, both the question and request bars offer 10 to 16 predefined messages, and the assert bar offers an open chat in which the TT can type messages. In the example in Figure 2, the TT is blocked from using the "speed" button, and its underlying information can be acquired only by questioning the agent.

Again, COLBAS measures CPS through two individual problem-solving processes (knowledge acquisition and knowledge application) and three collaboration dimensions (questioning, requesting, and asserting). According to MicroDYN, TTs score points for deriving the correct connections between variables in the model during the knowledge acquisition phase and for achieving goal states during the knowledge application phase in the COLBAS units (for details, see Krkovic et al., 2016). As the COLBAS knowledge acquisition phase is extended to group settings, the sum of collaborative actions is combined into three collaboration scores for the numbers of questions, assertions, and requests sent. Each collaborative action receives a score of 1, whereas no collaborative action receives a score of 0. As the assertion button involves open chat sessions, each assertion receives a score of 1, but the contents and the number of words are ignored (for further details, see Krkovic et al., 2016). COLBAS assesses CPS with six units that gradually increase in difficulty (Krkovic et al., 2016). The numbers of variables and/or relations between them are varied as in MicroDYN (Greiff et al., 2012, as cited in Krkovic et al., 2016). The difficulty is varied as in MicroDYN by increasing the number of variables and/or increasing the number of relations between them (Greiff et al., 2012).

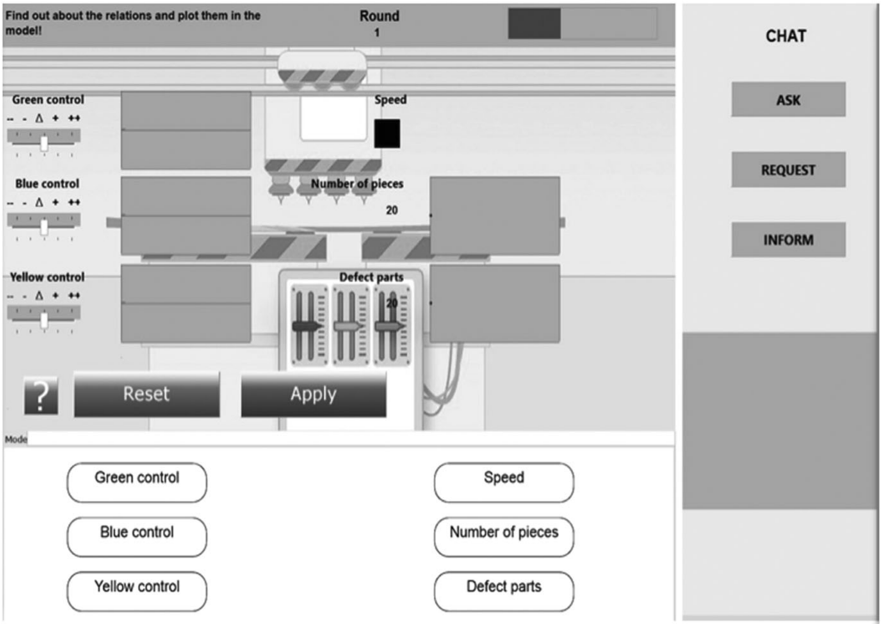

Figure 2. Screenshot of a knowledge acquisition phase in the COLBAS unit "Handball" (obtained from Krkovic et al., 2016). The left side of the screen shows the task space that the TT can use to explore the problem. The "speed" button cannot be explored by the TT, and its information can be acquired only by collaborating with the agent. The right side of the screen shows the communication options ask, request, and inform. Notably, the Krkovic et al. (2016) paper and this article use the term *question* instead of *ask* and the term *assert* instead of *inform* as contrastingly used within the COLBAS units. For further details on COLBAS, please see Krkovic et al. (2016). Reproduced with permission from *European Journal of Psychological Assessment*, Vol. 32(1): 52–60 ©2016 Hogrefe Publishing, www.hogrefe.com, DOI: 10.1027/1015-5759/a000329

Conceptual and Empirical Mapping of COLBAS Onto the PISA 2015 CPS Approach

The overarching PISA 2015 CPS approach allows for the conceptual embedding of COLBAS because the approaches overlap conceptually in specific aspects, with PISA 2015 being the more general and comprehensive approach. COLBAS measures the TTs' proficiency in exploring a stated problem and in acquiring knowledge to solve the problem through the knowledge acquisition phase. In comparison with COLBAS, PISA 2015 measured these proficiencies mainly by examining the individual problem-solving processes A (exploring and understanding), and B (representing and formulating) (for further information, see OECD, 2013). PISA 2015 and COLBAS conceptually overlap because the knowledge acquisition phase is integrated into the PISA 2015 individual problem-solving processes A and B (see Table 1). Not only do COLBAS and PISA 2015 conceptually overlap in knowledge acquisition, they also both assess the TTs' proficiency in applying acquired knowledge correctly to solve the stated problems. More specifically, COLBAS assesses this particular individual problem-solving proficiency in the units' knowledge application phases, whereas PISA 2015 did this through the individual problem-solving process C (planning and executing; see Table 1).

Not only do the individual problem-solving aspects of COLBAS conceptually overlap with PISA 2015; COLBAS collaboration dimensions can also be embedded in the wide overarching PISA 2015 CPS construct. The COLBAS collaboration dimensions questioning, asserting, and requesting were conceptualized on the basis of Searle's (1975; as cited in Krkovic et al., 2016) seminal speech act theory and are assumed to be manifested in the PISA 2015 collaboration dimensions (1) establishing and maintaining shared understanding, (2) taking appropriate action to solve the problem, and (3) establishing and maintaining team organization (see Table 1; cf. Krkovic et al., 2016). For instance, questioning is assumed to be required for successfully (1) establishing and maintaining a shared understanding (Krkovic et al., 2016). Assertions can be supportive in that they can "offer team members help and give them ideas about their own abilities" (Krkovic et al., 2016, p. 6), which is part of (2) taking appropriate action to solve the problem. Finally, requesting is assumed to be an integral part of (3) establishing and maintaining team organization (see Table 1). Therefore, the three COLBAS collaboration dimensions could also be embedded in the PISA 2015 collaboration dimensions. At this point, it is important to note that COLBAS treats individual problem-solving and collaboration dimensions as separate and not conjoint on the basis of empirical findings (cf. Krkovic et al., 2016). By contrast, PISA 2015 further crossed the individual problem-solving processes and the collaboration dimensions and conceptualized 12 individual CPS skills (see Table 1). Therefore, in the present research, we focused on the specific previously described COLBAS aspects that were integrated into the underlying PISA 2015 dimensions A–D and (1)–(3) and not so much on the separate 12 cells within the PISA framework.

Along with the conceptual embedding of COLBAS into PISA 2015 outlined above, the empirical embedding aimed to identify relations between the COLBAS aspects of the individual problem-solving and collaboration dimensions. This answers the question of whether differences in collaborator profiles also exist in H-A settings, as previously discussed. If so, identified CPS profiles not only add essential knowledge to CPS research, but also give first potential insights of CPS profiles in PISA 2015. CPS profiles can depict strengths and weaknesses in individual cognitive individual problem-solving aspects and collaboration aspects. Especially in a large-scale setting (such as in PISA 2015), different profiles of CPS behavior can enable a valuable international comparison between students, which may indicate differences in educational outcomes from different instructional teaching methods. Modern CPS teaching methods could be improved, if needed, and further support CPS learning in school. To reveal specific profiles of CPS behavior in students, we applied a model-based latent class cluster analysis to the COLBAS embedded aspects[4] in PISA 2015.

Model-Based Cluster Analysis

The model-based cluster analysis aimed to identify CPS behavior profiles in students on the basis of the aspects of COLBAS that were embedded in the PISA 2015

CPS approach. We posed three research questions. Research Question 1 addressed the type and number of profiles of CPS collaborators on the basis of the COLBAS data from Krkovic et al. (2016) (Research Question 1: How many distinct clusters of students with statistically distinguishable profiles of CPS behavior can be identified through students' COLBAS performance?). Research Question 2 addressed the validation of the profiles identified in Research Question 1 by including additional cognitive indicators (i.e., individual problem solving assessed by MicroDYN, and reasoning assessed by the Culture Fair Test (CFT) 20-R; Weiß, 2008) (Research Question 2: How do the identified profiles differ in cognitive indicators assessed by individual problem solving and reasoning?). Finally, Research Question 3 addressed the validation of profiles by including motivational indicators (i.e., collaboration motivation assessed by a question, which was asked once before each COLBAS unit, and test motivation assessed by statements, which were rated once after solving the six COLBAS tasks) (Research Question 3: How do the identified profiles differ in noncognitive indicators assessed by the collaboration motivation with the agent and overall test motivation?).

Participants[5]

A sample of $N = 509$ seventh graders from seven different secondary schools in Germany were randomly selected. Because secondary school is divided into a lower track (Hauptschule), a medium track (Realschule), and an upper track (Gymnasium) in Germany, the classes differed in school track assignment (for further information about the German school system, see Paulick, Watermann, & Nückles, 2013). Each class was assessed on two consecutive days during regular class time under the supervision of two trained test administrators. Participants completed a computer-based test battery of several tests and questionnaires that included various other measures (for details, see Krkovic et al., 2016). COLBAS, MicroDYN, CFT 20-R, collaboration motivation questions, and test motivation statements were parts of the test battery that are relevant to this study. MicroDYN, COLBAS, the collaboration motivation questions, and test motivation statements were assessed on the first test day, whereas the CFT-20R was assessed on the second test day. MicroDYN was assessed before COLBAS so the understanding of MicroDYN units was transferred to COLBAS; thus it was unlikely that adding collaboration to the individual problem-solving tasks (i.e., moving from MicroDYN to COLBAS) caused cognitive overload. For various reasons (e.g., technical difficulties), not all participants completed the COLBAS assessment (cf. Krkovic et al., 2016). A final convenience sample of $N = 481$ students was used for the analyses in this study.

Materials

Collaborative problem solving (CPS). COLBAS was used to assess CPS across a total of six units. COLBAS was introduced in an instructional video. The COLBAS assessment lasted approximately 45 minutes depending on the amount of collaboration (i.e., questioning, requesting, or asserting) with the agent. Individual problem solving in COLBAS (i.e., knowledge acquisition and knowledge application) was scored equivalently to the scoring used in MicroDYN (Krkovic et al., 2016). In the

COLBAS knowledge acquisition phase, individual problem solving in COLBAS was scored dichotomously such that TTs receive full credit (1) or no credit (0) for individual tasks. TTs scored full credit when they correctly drew connections between the input and output variables into the graphical model (cf. Greiff et al., 2012).[6] Collaboration was scored dichotomously, such that each TT received full credit (1) for each collaboration action (questioning, requesting, or asserting) or no credit (0) in each of the six individual tasks. As the assertion button involved open chat sessions, each assertion received a score of 1, but the content and the number of words were not analyzed (for further details, see Krkovic et al., 2016). Sum scores were calculated for each collaboration dimension (i.e., questioning, requesting, and asserting) across units. Further, in the knowledge application phase, TTs received full credit (1) when they used the model to achieve particular target values (for details, see also Wüstenberg et al., 2012), otherwise they received no credit (0).

Individual problem solving. Individual problem solving was assessed through MicroDYN (Greiff et al., 2012). MicroDYN tasks represent dynamic problem situations that consist of input and output variables. Students explore the relations between the input and output variables by acquiring knowledge in the first phase and applying the knowledge in the second phase. The MicroDYN tasks in the current study took approximately 45 minutes to complete (cf. Krkovic et al., 2016). Instructions and a practice task were run before the main assessment (for further details, see Krkovic et al., 2016). The main tasks were scored dichotomously such that TTs received full credit (1) when they drew the correct connections between the input and output variables in the knowledge acquisition phase and when they used the model to reach particular target values in the knowledge application phase. Otherwise participants received no credit (0).

Reasoning. The Culture Fair Test (CFT) 20-R was used to assess students' reasoning skills. The CFT 20-R was employed in two parts that both consisted of four subtests for assessing series, classification, matrices, and typologies. Only the raw data from the first test (56 items) was used in the analysis as suggested by Weiß (2008) for younger samples. The CFT assesses reasoning tasks on series, classification, matrices, and typologies. Reasoning was scored dichotomously such that TTs received full credit (1) or no credit (0) for individual tasks. The sum of points reflected the overall reasoning performance.

Test motivation. Answers to the statements "I have thoroughly worked on the units," "I did my best," "I made an effort," "I have endeavored to succeed in the units," and "I was concentrated" were requested once after participants solved the six COLBAS tasks. Test motivation statements were scored on a scale ranging from 1 (*disagree*) to 4 (*agree*), and were summed up to an overall test motivation score. Cronbach's alpha for the four items was .91, and therefore very good.

Collaboration motivation. The question "If you were able to choose, would you prefer working on the task alone or together with a computer-generated partner?" was asked once before each COLBAS unit. Collaboration motivation was scored dichotomously such that TTs received full credit (1) for answering that they preferred to collaborate with the partner or no credit (0) otherwise. The sum of the points reflected their level of collaboration motivation. Cronbach's alpha was .53, and therefore low.

Statistical Analysis

We identified distinguishable latent profiles of students' CPS behavior (Research Question 1) by computing a model-based cluster analysis in the Statistical Package for Social Sciences (SPSS v 19.0; IBM Corporation, Armonk, NY, United States) and *R* (v 2.15; R Foundation for Statistical Computing, Vienna, Austria). The *mclust* package (Fraley, Raftery, Murphy, & Scrucca, 2012) was chosen (a) because of the nature of our continuous data, (b) because we included a total of five COLBAS variables (i.e., knowledge acquisition, knowledge application, questioning, requesting, and asserting), and (c) because of its general power to estimate the best model fit according to different covariance structures and numbers of profiles (for further details, see Haughton, Legrand, & Woolford, 2009). The number of profiles and their patterns with regard to TTs' COLBAS performance were explored and determined on the basis of relative fit indices (Fraley & Raftery, 2002) and theoretical considerations as suggested by Geiser (2011). For Research Question 2, one-way between-subjects analyses of variance (ANOVAs) and pairwise Scheffé post hoc tests were computed to further validate cognitive differences between cluster profiles by including individual problem-solving and reasoning indicators. For Research Question 3, ANOVAs further validated the cluster profiles with noncognitive motivational indicators, which were identified by the collaboration motivation and test motivation.

Results

Research Question 1: Identifying the Number of Clusters and Their Profiles

Clusters were identified on the basis of their COLBAS performance (i.e., knowledge acquisition, knowledge application, questioning, requesting, and asserting). We identified the number of clusters (Research Question 1) by relying on relative fit indices (Fraley & Raftery, 2002; Geiser, 2011) as well as class attribution probabilities and interpretability with respect to theoretical considerations (Geiser, 2011). With respect to relative fit indices, a model-based cluster analysis provides best model fit by calculating a number of Gaussian models per cluster, whereby a Bayesian information criterion (BIC) is determined for each model (Fraley & Raftery, 2002). A cluster analysis assumes that the data integrated into the analysis were generated from a population that consisted of several subpopulations (Pieters, Roeyers, Rosseel, Van Waelvelde, & Desoete, 2015). Each subpopulation (or cluster) is modeled by Gaussian models of different types of multivariate normal distributions. The types of Gaussian models include spherical, ellipsoidal, or diagonal shapes as well as same or differently shaped distributions for each profile (for further details see Steinley & Brusco, 2011; Vermunt, 2011 as cited in Pieters et al., 2015). The best Gaussian model determined per cluster corresponds to the maximum BIC. Overall one- to nine-cluster solutions were calculated in our analysis. The three-cluster solution (BIC $= -6{,}319$) reported the highest BIC, closely followed by the eight-cluster solution (BIC $= -6{,}357$) and the seven-cluster solution (BIC $= -6{,}360$). Figure 3

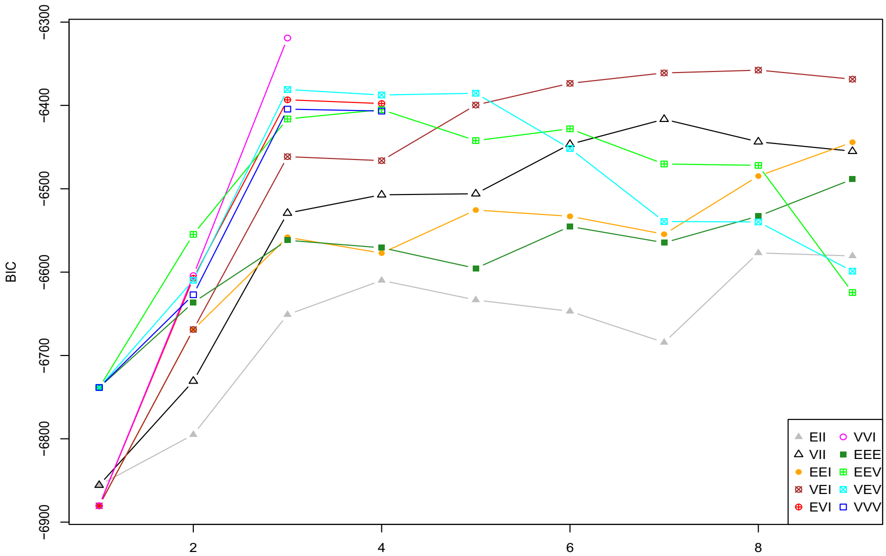

Figure 3. Bayesian information criterion (BIC) values for each Gaussian model considered in the model-based cluster analysis. The BIC plot shows each BIC value for each profile in which line graphs illustrate the different types of multivariate normal distributions integrated into the model per cluster. Each Gaussian model is illustrated with a different icon and a three-letter sequence. The letter sequence is a code for the geometric characteristics of volume, shape, and orientation. In the letter code, E means equal, V means varying across all clusters, and I specifies shape or orientation and is a special case of E. For a detailed description, please see Fraley and Raftery (2007).

shows a BIC plot that illustrates the comparison of each BIC value considered in the analyses for the one- to nine-cluster solutions. The best fit of the three-cluster solution was directly observable as an obvious peak.

The three-cluster solution was also supported if we considered class attribution probabilities, which is the probability of specific students belonging to the identified cluster in the model, and interpretability with respect to our theoretical assumptions. First, the three-cluster solution demonstrated attribution probabilities above .80 and indicated a sufficiently exact cluster allocation. More specifically, Cluster 1's attribution probability was .89, Cluster 2's was .88, and Cluster 3's was .94. Second, in terms of interpretability with respect to theoretical considerations, the three-cluster solution provided the most interpretable CPS profiles compared to the other eight- and seven-cluster solutions, also for the later validation of cognitive and noncognitive indicators. Thus, we considered the three-cluster solution to be the most appropriate for our study.[7] In this solution, three latent profiles were identified that differed in their COLBAS performance of (1) knowledge acquisition, (2) knowledge application, (3) questioning, (4) requesting, and (5) asserting. Based on their COLBAS performance, the three profiles were named as (1) passive low-performing (non-)collaborators (163 students), (2) active high-performing collaborators (171 students), and (3) compensating collaborators (147 students).

The passive low-performing (non-)collaborators showed a low level of individual problem-solving success (i.e., in knowledge acquisition and knowledge application) and the lowest levels of collaboration actions (i.e., in questioning, requesting, and asserting) in comparison with the other clusters. Based on this cluster's overall passiveness, this group was profiled as the "passive low-performing (non-)collaborators." In contrast to the passive low-performing (non-)collaborators, the active high-performing collaborators exhibited a cognitively strong performance by showing the significantly highest levels of individual problem-solving success across all clusters. Students assigned to this cluster collaborated with the agent at average levels compared with the other clusters. This profile was designated the "active high-performing collaborators" because these students had a cognitively strong overall performance and collaborated at least some extent. Finally, the last profile of the compensating collaborators performed comparably low on the individual problem-solving variables coupled with the highest collaboration actions (questioning, requesting, and asserting) as compared with the other clusters. This profile collaborated significantly more frequently than the other profiles. From this, we assume that this cluster compensated for relatively low cognitive performance by collaborating with the agent; thus, this profile was designated the "compensating collaborators." For a better illustration, Figure 4 provides a visual representation of the performance comparisons between the profiles. Please find the descriptive statistics of detailed performance comparisons between the profiles in Appendix 1 in the online Supporting Information.

Research Question 2: Cluster Membership and Cognitive Performance

In the second step, the three clusters were validated by investigating profile differences in two cognitive instruments (i.e., individual problem solving assessed by MicroDYN, and reasoning assessed by CFT 20-R) to substantiate the cluster profiles for individual problem-solving proficiency in COLBAS. In addition to means and standard deviations, pairwise Scheffé post hoc tests were computed to evaluate the cluster differences. The results indicated that the active high-performing collaborators exhibited significantly better cognitive performance on all three external indicators (MicroDYN knowledge acquisition, MicroDYN knowledge application, and reasoning). By contrast, the passive low-performing (non-)collaborators reported the lowest IPS performance across all clusters, but did not differ notably from the compensating collaborators. Figure 5 presents the cluster profiles, and Appendix 2 in the online Supporting Information reports a detailed comparison of the class differences.

Research Question 3: Cluster Membership and Motivation

In a third step, we investigated the collaboration motivation with the agent as an indicator of the motivation to actively question, assert, and request in the COLBAS units. In addition, we measured students' test motivation. Participates' collaboration motivation scores were summed into an overall collaboration motivation score, and individual test motivation scores were summed into an overall test motivation score. The active high-performing collaborators reported the lowest collaboration

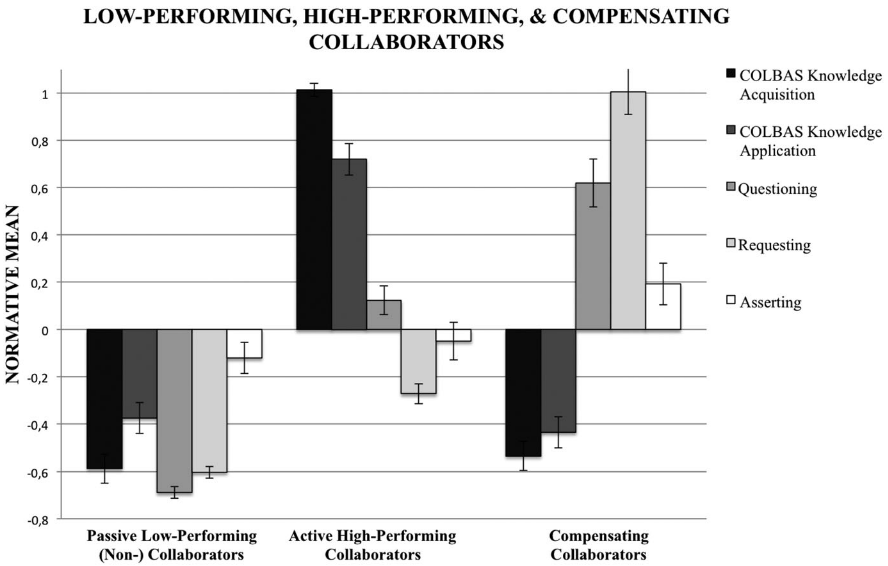

Figure 4. Bar graph with standard error bars (*SE*) representing cluster profiles for the passive low-performing (non-) collaborators (Cluster 1), active high-performing collaborators (Cluster 2), and compensating collaborators (Cluster 3). Standardized *z*-scores in the full sample were generated for each variable to illustrate the relative differences between the clusters. Values for each variable at the zero mean (0) are classified as average performance. Values below the zero mean indicate lower performance, and values above the zero mean indicate a higher performance than average. The *x*-axis displays the performance of the different clusters in the five clustering variables. The *y*-axis displays the mean scores for drawing the correct model (knowledge acquisition), achieving a specific goal (knowledge application), and communicating with the agent (questioning, requesting, and asserting).

motivation with others; however, they reported average levels of collaborative actions (i.e., questioning, asserting, and requesting; see Appendix 1, Figure 5, in the online Supporting Information) compared with the other two collaborator profiles. In addition, they reported the highest test motivation compared to the other profiles (see Appendix 1 in the online Supporting Information). Great differences in test motivation and collaboration motivation indicated that students were more focused on individual performance than collaboration, or vice versa. The compensating collaborators reported a significantly higher collaboration motivation compared to the passive low-performing (non-)collaborators and active high-performing collaborators. The passive low-performing (non-)collaborators chose to work with the agent instead of working alone on average levels; however, they remained relatively passive in all collaborative actions (i.e., questioning, asserting, and requesting) compared with the active high-performing collaborators and compensating collaborators (see Appendix 1, Figure 5, in the online Supporting Information). They also reported the lowest test motivation across clusters.

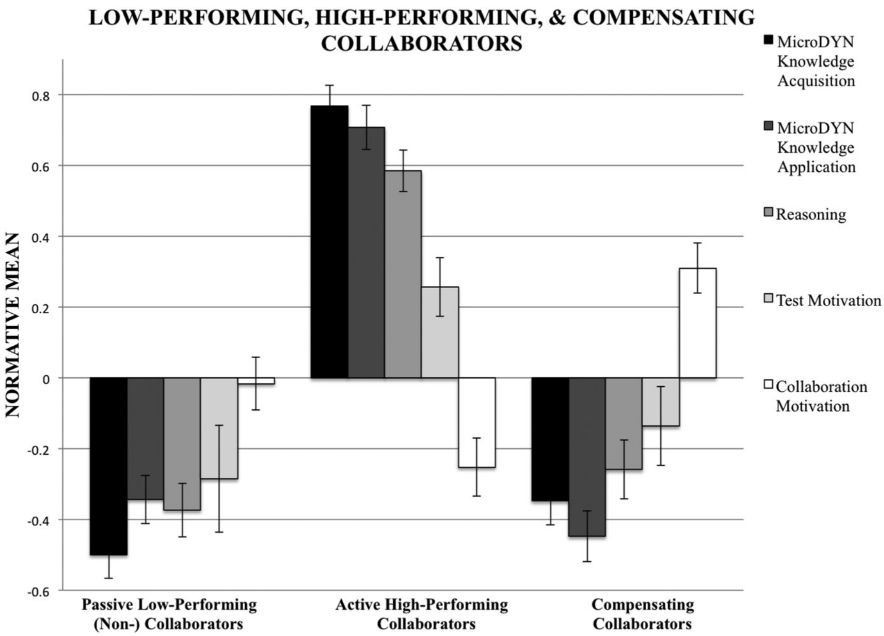

Figure 5. Bar graph with standard error bars (*SE*) representing cluster profiles for the passive low-performing (non-)collaborators (Cluster 1), active high-performing collaborators (Cluster 2), and compensating collaborators (Cluster 3). Standardized *z*-scores in the full sample were generated for each variable to illustrate the relative differences between the clusters. The values for each variable at the zero mean (0) are classified as average performance. The values below the zero mean indicate lower performance, and the values above the zero mean indicate a higher performance than average. The *x*-axis displays the performance of the different clusters in the five clustering variables. The *y*-axis displays the mean scores for the overall reasoning score, collaboration motivation with the agent, and test motivation.

Discussion

The aim of this article was to review the PISA 2015 H-A approach and to theoretically map COLBAS onto this overarching approach. This enabled the empirical identification of CPS profiles by conducting a cluster analysis with additional cognitive and motivational indicators. First, we theoretically revisited the PISA 2015 CPS assessment approach. Second, the PISA 2015 CPS assessment allowed for the conceptual embedding of existing H-A instruments for CPS assessment such as COLBAS. We conceptually embedded the COLBAS variables of knowledge acquisition, knowledge application, questioning, requesting, and asserting in the overarching PISA 2015 CPS dimensions A–C for individual problem solving, and (1)–(3) for the collaboration dimensions. We applied an empirical approach of model-based cluster analysis to examine the specific aspects of COLBAS that were embedded in PISA 2015 in order to provide some first potential insights into CPS profiles in PISA 2015 as a complement for later reporting. The latent cluster analysis identified the three profiles of passive low-performing (non-)collaborators, active high-performing collaborators, and compensating collaborators. Investigating profile differences in cognitive indicators (i.e., individual problem solving and reasoning) and motivational indicators (i.e., test motivation and collaboration motivation) further validated the clusters. It should be noted that the test motivation is not included in the PISA 2015 CPS framework. However, understanding participants' levels of test motivation served the purpose of validating participants' collaboration motivation.

The passive low-performing (non-)collaborators remained inactive in their cognitive performance and collaborative actions and reported low overall motivation and average collaboration motivation compared with the active high-performing and compensating collaborators. By contrast, the active high-performing collaborators exhibited strong cognitive performances and test motivation, while collaborating with the agent despite having the lowest collaboration motivation. The compensating collaborators exhibited low cognitive performance but had the highest collaboration motivation across all profiles. We assume that this cluster collaborated strongly to compensate for their low individual problem-solving skills. Through their high level of requesting, TTs in this group may have tried to obtain direct solutions from the agent without the need to show individually high levels of performance in the problem-solving part. Interestingly, the compensating and passive low-performing (non-)collaborators showed similar performance patterns in their individual problem-solving performance in the COLBAS, MicroDYN, and reasoning task. The results for the compensating and active high-performing collaborators are similar with profiles found in H-H settings; Chung, O'Neil, and Herl's (1999) study found low-performing groups sending predefined messages, gaining less knowledge from conversations, and exhibiting a lower performance, and vice versa. Also, Rummel, Mullins, and Spada (2012, as cited in Krkovic et al., 2016) suggest that high levels of interaction do not guarantee success on CPS units because the interactions may be inefficient and not lead to an actual solution of the problem. This was the case with the compensating collaborators, who collaborated strongly but reported low problem-solving performances. Overall, these studies were in line with our results on profiles in our sample. However, because research in this field is scarce, further work should be conducted to replicate and validate our profiles in other samples.

Considering the three clusters obtained in this study as potential CPS profiles in PISA 2015, the depiction of strengths and weaknesses in specific individual cognitive problem-solving aspects and collaboration aspects could provide valuable insights into students' current CPS performance. Especially in an international large-scale setting such as PISA 2015, these insights provide information on students' current cognitive and collaborative performance as educational outcomes of CPS learning at school that is gained from different instructional teaching methods. Identified weaknesses in CPS behavior can provide fundamental insights that in turn can be used to improve CPS teaching methods in educational systems. Considering that cluster analyses have these strengths, we advocate cluster analyses as appropriate for the PISA 2015 CPS data in order to identify further information on students' CPS behavior, such as particular CPS skills, in addition to their overall CPS scores as determined in PISA 2015. It is important to note that PISA 2015 treated CPS as a combination of individual problem-solving processes and collaboration dimensions, and thus our potential CPS profiles are applicable only for the overarching PISA 2015 individual problem-solving and collaboration dimensions. Therefore, cluster calculations on the PISA 2015 results—meaning the classification of subgroups of CPS behavior on the basis of the 12 CPS variables and not five as in COLBAS—may be able to identify profiles on an even more fine-grained level. These cluster profiles are predicted to include additional indicators that are not measured in COLBAS (i.e., besides questioning, asserting, and requesting) that might also play an important role in collaboration in terms of the PISA 2015 theoretical framework. For example, questioning, which was embedded in (1) establishing and maintaining a shared understanding in a group, is only one aspect of collaborative behavior that continuously creates a shared understanding. Other parts of problem-solving processes that are crossed with (1), such as negotiating or monitoring, should also be integrated. Therefore, cluster calculations with PISA 2015 data could identify rich profiles of CPS behavior providing information on their performance in specific subskills. Considering that the 12 CPS skills in PISA 2015 were conceptualized by giving more weight to the collaboration dimensions than the individual problem-solving processes (for further details, see OECD, 2013, p. 24), applying the cluster analyses to the PISA 2015 results may create profiles that reflect more noncognitive collaboration skills than individual problem-solving skills. Besides these additional opportunities cluster analysis can provide, further opportunities, such as researching specific information on age, country, language, culture, and cluster membership from multiple countries, can be researched by the application of cluster analyses to the PISA 2015 results. In PISA 2015, clusters can be identified on a higher number of variables compared to this study, and therefore provide rich information on CPS behavior differences among students.

Conclusion

This article reviewed the PISA 2015 CPS approach and, in addition, empirically measured specific aspects of the PISA 2015 approach by embedding the COLBAS approach in the PISA 2015 framework and applying a model-based cluster analysis. Our analysis identified three profiles of CPS behavior: passive low-performing (non-)collaborators, active high-performing collaborators, and compensating collaborators, each differing in cognitive performance, collaboration, and motivation. The identified CPS profiles can serve as a good starting point for PISA 2015. However, to identify profiles of CPS behavior in PISA 2015 and highlight students' current major strengths and weaknesses in CPS across different educational systems, cluster analyses could be applied after the PISA 2015 CPS data are published in late 2017, complementing the standard quantitative reporting usually employed in PISA.

Notes

[1]This article uses the acronym CPS to indicate collaborative problem solving. The acronym CPS also stands for complex problem solving in the scientific literature (Greiff, Wüstenberg, & Funke, 2012).

[2]This article uses the term *individual problem solving*; PISA 2012 used the term *creative problem solving* (OECD, 2012).

[3]The embedded COLBAS aspects (in bold) are relevant for the later cluster analysis and detailed information will be provided in the later paragraphs.

[4]The cluster analysis was a reanalysis of the raw data generated by Krkovic et al. (2016).

[5]We present here a reanalysis of a preexisting data set generated by COLBAS (Krkovic et al., 2016).

[6]Technical problems resulted in the failure to capture one knowledge acquisition indicator of one of the COLBAS units during testing.

[7]We additionally calculated a latent class analysis (LCA) in MPlus Version 7.0 (Muthén & Muthén, 2012) to validate our results. The analysis also classified our sample into the three latent classes as classified in *R*. We preferred the cluster analysis in *R* using mclust for the main analyses in this article so that we could achieve cluster results on a more fine-grained level.

Acknowledgments

This research was funded by a grant from the Fonds National de la Recherche Luxembourg (ATTRACT "ASKI21") awarded to Samuel Greiff. We thank Katarina Krkovic and Julia Rudolph for their invaluable help with this article including their contribution to data collection and data preparation.

References

Chung, G. K. W. K., O'Neil, H. F., & Herl, H. E. (1999). The use of computer-based collaborative knowledge mapping to measure team processes and team outcomes. *Computers in Human Behavior*, *15*, 463–493.

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, *97*(458), 611–631.

Fraley, C., & Raftery, A. E. (2007). Model-based methods of classification: Using mclust software in chemometrics. *Journal of Statistical Software*, *18*, 1–13.

Fraley, C., Raftery, A. E., Murphy, B., & Scrucca, L. (2012) *mclust version 4 for R: Normal mixture modeling for model-based clustering, classification, and density estimation* (Technical report No. 597). Department of Statistics, University of Washington.

Geiser, C. (2011). *Datenanalyse mit MPlus*. Wiesbaden, Germany: Verlag für Sozialwissenschaften.

Greiff, S. (2012). From interactive to collaborative problem solving: Current issues in the Programme for International Student Assessment. *Review of Psychology*, *19*(2), 111–121.

Greiff, S., Wüstenberg, S., & Funke, J. (2012). Dynamic problem solving a new assessment perspective. *Applied Psychological Measurement*, *36*, 189–213.

Haughton, D., Legrand, P., & Woolford, S. (2009). Review of three latent class cluster analysis packages: Latent Gold, poLCA, and mclust. *American Statistician*, *63*(1), 81–91.

Krkovic, K., Wüstenberg, S., & Greiff, S. (2016). Assessing collaborative behaviour in students: An experiment-based assessment approach. *European Journal of Psychological Assessment*, *32*, 52–60.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

OECD. (2012). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris, France: OECD.

OECD. (2013). *PISA 2015 draft collaborative problem solving framework*. Paris: OECD.

OECD. (2015a). *About PISA*. Retrieved from http://www.oecd.org/pisa/aboutpisa/

OECD. (2015b). *PISA 2015 released field trial: Cognitive items*. Paris, France: OECD.

O'Neil, H. F., Jr., Chuang, S. H., & Baker, E. L. (2010), Computer-based feedback for computer-based collaborative problem solving. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 261–279). New York, NY: Springer.

Paulick, I., Watermann, R., & Nückles, M. (2013). Achievement goals and school achievement: The transition to different school tracks in secondary school. *Contemporary Educational Psychology*, *38*(1), 75–86.

Pieters, S., Roeyers, H., Rosseel, Y., Van Waelvelde, H., & Desoete, A. (2015). Identifying subtypes among children with developmental coordination disorder and mathematical Learning Disabilities, using model-based clustering. *Journal of Learning Disabilities*, *48*(1), 83–95.

Rosen, Y., & Tager, M. (2013). *Computer-based assessment of collaborative problem solving skills: Human-to-agent versus human-to-human approach*. Boston, MA: Pearson Education.

von Davier, A. A., & Halpin, P. F. (2013). *Collaborative problem solving and the assessment of cognitive skills: Psychometric considerations* (ETS Research Report Series, 2013(2), i–36). Princeton, NJ: Educational Testing Service.

Weiß, R. H. (2008). *Grundintelligenztest Skala 2 - Revision - (CFT 20-R) [Culture Fair Intelligence Test, Scale 2 - Revision]*. Göttingen, Germany: Hogrefe.

Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving: More than reasoning? *Intelligence*, *40*, 1–14.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site.

Appendix 1. Means, Standard Deviations (in Brackets) and Pairwise Scheffé—Comparisons Between the Three Clusters on COLBAS Performance

Appendix 2. Means, Standard Deviations (in Brackets), and Pairwise Scheffé—Comparisons Between the Three Clusters on MicroDYN Performance, Reasoning, and Motivational Indicators

Authors

KATHARINA HERBORN is Doctoral Candidate, Computer-Based Assessment Research Area, Institute of Cognitive Science and Assessment, University of Luxembourg, 11, Porte des Sciences, L-4366 Esch/Alzette; katharina.herborn@uni.lu. Her primary research interests include the assessment of collaborative problem solving.

MAIDA MUSTAFIĆ is Research Associate, Computer-Based Assessment Research Area, Institute of Cognitive Science and Assessment, University of Luxembourg, 11, Porte des Sciences, L-4366 Esch/Alzette; maida.mustafic@uni.lu. Her primary research interests include lifelong development.

SAMUEL GREIFF is Professor, Computer-Based Assessment Research Area, Institute of Cognitive Science and Assessment, University of Luxembourg, 11, Porte des Sciences, L-4366 Esch/Alzette; samuel.greiff@uni.lu. His primary research interests include educational assessment, problem solving, and multivariate statistics.