Fischer, Jessica; He, Jia; Klieme, Eckhard

The structure of teaching practices across countries. A combination of factor analysis and network analysis

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:
formally and content revised edition of the original source in:
Studies in educational evaluation 65 (2020) 100861, 55 S.

The Structure of Teaching Practices across Countries: A Combination of Factor Analysis and Network Analysis

Jessica Fischer*
DIPF | Leibniz Institute for Research and Information in Education

Jia He*
DIPF | Leibniz Institute for Research and Information in Education
Tilburg University, the Netherlands

Eckhard Klieme
DIPF | Leibniz Institute for Research and Information in Education

* These authors contributed equally to this work.

Funding

This work was partially supported by the Marie Sklodowska-Curie Individual Fellowship European program [grant number 748788]

Cite

Fischer, J., He, J., & Klieme, E. (2020). The structure of teaching practices across countries: A combination of factor analysis and network analysis. *Studies in Educational Evaluation, 65*. https://doi.org/10.1016/j.stueduc.2020.100861

Abstract

Teaching practices are pivotal for student learning. Due to pedagogical traditions and national cultures, the structure of teaching practices may differ across countries. This study investigates the structure of teaching practices across 12 countries grouped into four major linguistic/cultural clusters. First, factor analysis is applied to investigate if the theoretical distinction between teacher-directed and student-centred practices is generalizable across countries. Then, network analysis is used to explore how individual classroom assessment practices relate to either teacher-directed or student-centred practices. Main findings include that: (1) teacher-directed and student-centred practices are two distinct factors across countries; (2) the overall structure and connectivity of teaching practices differs across countries, with smaller differences within linguistic/cultural clusters; and (3) assessment practices with the aim to structure and guide learning strongly relate to teacher-directed practices, whereas assessment practices with the aim to individualize instruction more relate to student-centred practices. We discuss the global patterning and implications.

Keywords: teaching practices, classroom assessment, factor analysis, network analysis, PISA, cross-cultural

1 Introduction

Across the world, teacher's instructional practice has been recognized to be one of the most important factors influencing student learning outcomes (Hattie, 2009). Teaching practices can be seen as a major part of classroom instruction and in contrast to other factors relevant for student learning (e.g., the student's socio-economic background) they are more readily modifiable and, thus, can be subjected to targeted interventions (Vieluf, Kaplan, Klieme, & Bayer, 2012). In the last decades, international research often discussed two approaches to teaching, deemed opposite to each other, based on philosophies of education: teacher-directed and student-centred teaching practices (Tobias & Duffy, 2010). However, it has been argued that these theoretical conceptualizations do not account for the complex nature of teaching practices. First, there is no single best way of teaching; instead teachers are required to combine various strategies depending on the context, class, and students (Echazarra, Salinas, Méndez, Denis, & Rech, 2016). Research to identify how various practices relate to each other and what is the most beneficial mix is still scarce. Second, teaching can be regarded as cultural activity and is not generalizable across countries (Stigler & Hiebert, 1999). National culture and pedagogical tradition interactively influence approaches to teaching, leading to differences in frequency and combination of teaching practices, and consequently challenging the assumption of a universal structure of teaching practices. Additionally, the international debate recently considers classroom assessment (e.g., providing feedback) as part of instructional practices that teachers implement in the classroom (OECD, 2013). Yet, it remains unclear how classroom assessment relates to teacher-directed and student-centred practices. In order to tailor targeted interventions to promote high-quality teaching in a culturally sensitive way, a comprehensive understanding of the structure and co-occurrence of teaching practices across countries is indispensable.

In this study, we aim to shed light on the structure and co-occurrence of teaching practices across countries in two steps. First, we check if the theoretical distinction between teacher-directed and student-centred practices is empirically supported across countries. Secondly, we investigate how classroom assessment practices (which were rarely simultaneously tested with teacher-directed and student-centred practices in empirical studies) integrate into the broader framework of teaching practices with an exploratory approach. More precisely, we investigate if individual classroom assessment practices differently relate to either teacher-directed or student-centred practices. Given that high-quality teaching requires teachers to combine diverse practices to foster student learning, we propose to consider direct relationships between individual practices beyond focusing on the shared underlying factors. Thus, we complement conventional factor analysis with psychological network analysis. Network analysis models direct interactions among individual practices and helps visualize the "ecosystem" (e.g., conditional co-occurrence) of teaching practices. It helps us 1) to illustrate the conditional co-occurrence of practices and compare the patterning across countries, 2) to account for the interdependency without reducing the related practices to a single construct score, as has been done in studies on teaching practices (see e.g., OCED, 2019), and 3) to provide a foundation to further explore overarching teaching quality dimensions in the future.

In the following, we first review the framework of teaching practices and recent developments, and highlight the importance of a cross-cultural perspective on teaching practices. Thereafter, we address the challenges of measuring and analysing teaching practices across countries.

1.1 Teacher-directed versus student-centred teaching practices

In the 20th century, educational theories have undergone significant developments. Influenced by the *behaviourism* in the United States (Carroll, 1963), and the German schools *Reform pedagogy* (e.g., by Peter Petersen) and *Gestalt psychology* (Duncker & Lees, 1945) in Western Europe, instructionist and constructivist theories of learning have emerged. Rooted in Western countries, both frameworks are increasingly influential outside North America and Western Europe. Instructionists such as Rosenshine (1976) characterize a traditional and teacher-directed approach to instruction with an information-processing view of learning. In contrast, constructivism, based on work of Vygotsky (1978), Dewey (1929), and Piaget (1952), promotes an alternative approach with the focus on the learner and learning context (Tobias & Duffy, 2010). These two dominant frameworks inspire approaches to designs of instruction to date, yielding the development and application of different teaching practices.

Direct instruction advocates the use of *teacher-directed* practices that aim to provide a well-structured and effective learning environment (Caro, Lenkeit, & Kyriakides, 2016). The teacher is the transmitter of knowledge and controls learning processes in the classroom. Besides delivering information, the teacher plans lessons in advance and structures the presentation of ideas in class. Guided by the teacher, students can participate during instruction, for instance, through answering the teacher's questions, posing own questions to the teacher, or reproducing received information (Mostafa, Echazarra, & Guillo, 2018). The advantage of teacher-directed practices is the emphasis on well-structured lessons, wider subject coverage, and a better preparation for standardized tests (Ormrod, 2012). Yet, the rather passive role of students can lead to a decline in motivation and positive attitudes towards the subject (Echazarra et al., 2016).

Student-centred teaching practices based on constructivism foster students' active engagement in their learning processes and promote a self-directed construction of knowledge. This can be facilitated by assigning activities that involve students in planning classroom activities, promoting discussions among students themselves and with the teacher, or by creating cooperative learning environments (e.g., small group work) while taking individual students' needs into account (e.g., achievement levels). The role of the teacher is to support and guide the learning processes. Student-centred practices are supposed to foster communication skills and collaborations, encourage students to direct their learning, and develop interest in a subject. Yet, student-centred practices are harder to implement and are often criticized to lack guidance for the learner, overtaxing working memory (Tobias & Duffy, 2010), and risking the development of incorrect knowledge (Mostafa et al., 2018).

In reality, teaching is often a combination of diverse practices (Klieme, 2020). In line with this reasoning, educational effectiveness research criticises this theoretical distinction to be insufficient to fully benefit student learning. Instead, the complementary application of both teacher-directed and student-centred practices is often seen to be more effective (Vieluf et al., 2012). Thus, the co-occurrence of teaching practices should be considered when conceptualizing and analysing teaching practices.

1.2 Classroom assessment practices

Teacher-directed and student-centred practices stem from learning theories dating back to decades ago (Richardson, 2003). More recently, classroom assessment, as an additional element of teaching practices, has garnered much attention (OECD, 2013) and is considered one of the most powerful teaching practice for quality management and the improvement of student learning outcomes (Klieme, 2020). For instance, the Program for International Student Assessment (PISA),

a triennial large-scale assessment of 15-year-old students in dozens of countries, operationalizes classroom assessment as a specific dimension of teaching practices (in addition to teacher-directed and student-centred practices) (OECD, 2014).

Classroom assessment practices are used to evaluate students' knowledge and progress (Coombs, DeLuca, LaPointe-McEwan, & Chalas, 2018). Depending on the standardization and purpose of the assessment, teachers possess a repertoire of tools to gather evidence about their students' progress and ideas (Harlen, 2007; Kippers et al., 2018). These classroom assessment practices can serve the purpose of *summarizing* the achievement of students or the *formative* purpose of improving teaching and learning on an ongoing assessment basis (e.g., discussion or oral examination) (Black & Wiliam, 2009). Drawing on Ramaprasad's theory (1983), formative assessment includes three steps: 1) identifying the current learning state, 2) establishing learning goals, and lastly 3) defining the steps that are needed to achieve the learning goals. An integral part of formative assessment is feedback. Astin and colleagues (1996) suggest that assessment is most effective when diverse methods are implemented complementary. Especially formative assessment combined with feedback has been shown to be a powerful tool to improve student achievement and motivation (Harlen & Deakin-Crick, 2002; Hattie, 2009).

Echazarra and colleagues (2016) placed classroom assessment between both traditional (teacher-directed) and modern (student-centred) ends of a teaching practice scale, yet there is rarely empirical evidence supporting this classification. Moreover, as assessment practices have to be applied by teachers with either teacher-directed or student-centred approaches in order to identify the students' learning state and progress, the question remains if and how different assessment practices are incorporated into different approaches to teaching.

1.3 Teaching practices across countries

Both the constructivist and instructionist framework were mainly developed and empirically tested in Western countries, and they might not be easily transferable to other cultures. While policy-makers across the world have the consensus on the importance of promoting high-quality teaching, they may have a different understanding of the structure of teaching practices and the notions of good practices. Praetorius and colleagues (2018) surveyed educational researchers from different countries regarding what constitutes good practices in their respective country, and they found substantial cross-country differences with regard to the categorization of good practices depending on pedagogical traditions and national cultures. For instance, practices promoting deep thinking, students' autonomy, and adaptive teaching were especially important in South-American countries, whereas East-Asian countries mostly valued practices ensuring well-structured lessons and independent thinking. German researchers defined feedback, addressing student errors, orderly managing the class, cognitive activation, and social-emotional support as important practices in their country. Furthermore, the effects of teaching practices on learning outcomes may be moderated by differences in educational systems or economic and cultural factors. For instance, Fuller and Clarke (1994) argued that student-centred practices promoting an active engagement of students during instruction are incompatible with strong hierarchical structures in countries valuing power distance. Likewise, McCormick and Alavi (2004) postulated that practices promoting teachers' critical reflection and inquiry might be less effective in collectivist countries, where criticism is communicated more indirectly than in Western countries. Consequently, the prevalent approaches to instruction and co-occurrence of teaching practices are likely to differ across countries. In a similar vein, cross-cultural research reported different frequencies of teacher-

directed, student-centred, and classroom assessment practices across countries, yielding different teaching practice profiles (see e.g., Echazarra et al., 2016).

It is important to emphasize, that teaching practice is difficult to generalize; instead it can be described as "cultural activity" (see Stigler & Hiebert, 1999) as it exhibits vast cross-cultural differences, not only in quantity, but also in quality, processes, and effectiveness. Thus, it is vital to consider context-specificity when measuring and analysing teaching practices cross-culturally (Vieluf et al., 2012).

1.4 Measuring and analysing teaching practices across countries

The dynamic model of educational effectiveness by Creemers and Kyriakides (2006) proposes to refine the measurement of teaching and learning constructs along multiple dimensions including frequency (i.e., the quantity that an activity is present in a system, school, or classroom), focus (i.e., the specificity and purpose of an activity), stage (i.e., the phase of an activity, with the assumption that the activity needs to take place for a long period of time to accumulate effects on student learning), quality (i.e., properties of the activity and its optimal use), and differentiation (i.e., the extent to which the activity is implemented for and has impact on all subjects in the same way). These measurement dimensions capture not only quantity but also quality and processes. Methodologically, teaching practices can be assessed with self-reports in surveys (from teachers and/or students) (e.g., OECD, 2015) and behavioural coding in video studies (e.g., Jacobs, Hollingsworth, & Givvin, 2007). In large-scale educational assessment, where many countries are compared (quantitatively), survey-based measurement is more frequently applied than behavioural coding, as it has the advantage of easy and cost-effective implementation to achieve sufficient sample sizes/power and to draw inferences about populations. Currently, large-scale surveys have a strong focus on the frequency dimension, and teaching practices are mostly assessed through

students' perceptions and experiences. For example, PISA asks students, the recipients of teaching practices, to report on the frequency of teachers' practices in classroom settings. Although such reporting does not tap into the quality or effectiveness of teaching practices, it provides data to test the structure and co-occurrence of teaching practices. However, as subjective teacher- or student-reports can be vulnerable to measurement bias, data quality and comparability across countries have to be tested (e.g., Vieluf, Kunter, & van de Vijver, 2013).

1.4.1 Multigroup confirmatory factor analysis (MGCFA)

Various psychometric tools are available to uncover the structure and metrics of self-reported teaching practices across countries. A conventional, rigorous approach involves multigroup confirmatory factor analysis (MGCFA). The common assumption with factor analysis is that items are indicators of latent factors and responses on items are "caused" by the latent factors. MGCFA provides a theory-driven approach (with a known factor structure) to examine a series of nested models across countries. These models include configural (i.e., the same configuration of zero and nonzero loadings of items on latent factors across countries), metric (i.e., the same factor loadings across countries), and scalar invariance (i.e., the same factor loadings and item intercepts across countries). Implications are attached with each level of invariance reached: configural invariance serves as a basis for any cross-country comparison, metric invariance allows valid comparisons of the unstandardized associations of constructs across countries (e.g., correlations between teaching practices and student outcomes), and only with scalar invariance can scale scores be compared across countries (i.e., means of teaching practices can be compared across countries, see Van de Vijver & Leung, 1997).

1.4.2 Network analysis

TALIS 2018 demonstrated that the different kinds of teaching practices are related (OECD, 2019) and, thus, it is important to consider the interdependency of teaching practices, without reducing them to a single "teaching practice score". Network analysis offers a novel perspective to gain insight into the co-occurrence (direct interactions) of observed indicators (Epskamp, Rhemtulla, & Borsboom, 2017) and helps us to understand how teaching practices are loosely or firmly related to each other as a system. It has been applied and is increasingly popular in personality research (see e.g., Costantini et al., 2015), research on political attitudes (see e.g., Dalege, Borsboom, van Harreveld, & Maas, 2018), and educational research (Abacioglu, Isvoranu, Verkuyten, Thijs, & Epskamp, 2019; Sachisthal, Jansen, Peetsma, Dalege, van der Maas, & Raijmakers, 2019). In contrast to factor analytic models, network analysis shifts the focus from the common shared variance to the variance between indicators (e.g., individual practices). Instead of assuming a common latent factor (e.g., "extraversion"), indicators (e.g., "I like to party", "I have a lot of friends") in a network are considered to mutually, directly affect each other – a change of one indicator leads to changes in the other connected indicator (e.g., by going to more parties, people meet more potential friends. And having more friends leads to more invitations to parties, see Costantini et al., 2015). Thus, indicators are part of the construct, instead of being measures of it (Sachisthal et al., 2019). The set of indicators (=*nodes*) and their interactions (=*edges*, representing unknown statistical relationships between two nodes) are visualized as a network, and magnitude (strength) and direction (positive versus negative) of pairwise interactions of indicators can be interpreted. Thus, network analysis illustrates if two teaching practices tend to co-occur frequently (positive relation) or not (absence of relations or negative relation) as well as the strength of their relation (if they are firmly or loosely related).

The function of indicators (i.e., nodes) within a network can be studied by examining their importance within the network (strength-centrality) or the structure and connectivity of the network. Nodes with higher values of *strength-centrality* influence other nodes more strongly than less central nodes, and thus, are the optimal starting point for targeted interventions or processes (Costantini et al., 2015). For teaching-practices, we view practices with the highest strength-centrality as the binding teaching practices (easily assigned with other practices) in the network. The *overall network structure* indicates the patterning of unique interactions between indicators in the network, and the *global connectivity* indicates the extent to which these indicators are connected (i.e., the extent to which teaching practices frequently co-occur) (Christensen, Kenett, Aste, Silvia, & Kwapil, 2018; Epskamp & Fried, 2018). The structure of a set of indicators and their overall connectivity can be compared across networks for different groups (i.e., different countries) by performing a network comparison test (NCT; van Borkulo, Epskamp, & Millner, 2016).

1.4.3 Combining MGCFA and network analysis

Network analysis can complement MGCFA in several ways. First, although no latent factor is assumed or pursued in network analysis, clusters of indicators linked by strong edges may be indicative of latent factors underlying these indicators, making network analysis a diagnostic tool to explore the dimensionality of constructs. Secondly, network analysis focuses on the intricate interactions, providing a differing and additional nuanced look at the dynamics among indicators as a system. Thirdly, comparisons based on network analysis do not require scalar invariance to be achieved. MGCFA aims to test whether individual or country means on the latent constructs can be compared validly (scalar invariance) and when scalar invariance is not tenable, the validity of further analysis on country means is not warranted (e.g., Vieluf et al., 2013). Yet, with many

different countries included in a study, the shared core of a construct becomes smaller, making it nearly impossible to achieve scalar invariance (analysis paradox, see van de Vijver, 2018). With network analysis, meaningful relations among items can be compared without pursing scalar measurement invariance. It has to be noted that measurement bias due to translation errors or different interpretations of the item content across groups (i.e., different countries) can nevertheless threat the validity and comparability of analysis results that are based on item responses (including network analysis).

1.5 The current study

We have summarized ongoing developments in the international debate on teaching practices and pointed out the possibility of country-specific structures and metrics in teaching practices, which cast doubts on the generalisability of a fixed structure of teaching practices. Nonetheless, empirical evidence on the structure of teaching practices across countries, with a focus on co-occurrence of teacher-directed, student-centred, and an addition of classroom assessment practices, is lacking. Thus, we formulate the following hypothesis and research question to explore the structure of teaching practices across countries.

Teacher-directed and student-centred teaching practices are based on well-founded theoretical approaches to instruction, particularly in Western countries. Moreover, both approaches have been operationalized and assessed in educational large-scale studies in dozens of countries (e.g., PISA, see OECD, 2014 and TALIS, see OECD, 2013), and they have guided designs to instruction across educational systems (e.g., Chile, see Zurita & Nussbaum, 2004 and Turkey, see Isikoglu, Basturk, & Karaca, 2009). Consequently, we expect to empirically identify the theoretically derived two distinct factors (i.e., teacher-directed versus student-centred practices) across countries (configural and metric invariance) (*Hypothesis 1)*. However, given the

considerable influence exerted by culture and pedagogical traditions on the design of instruction, the likelihood that individuals from different countries understand and respond to this set of indicators in exactly the same way is low. Furthermore, unintended differences between cultures (e.g., how respondents make use of the response scale in the frequency-based measures) may further endanger cross-cultural comparability of scale scores on teaching practices obtained in large-scale educational surveys. Thus, we expect cross-country differences in item intercepts of teacher-directed and student-centred practices (no scalar invariance, *Hypothesis 2*), challenging the full scalar comparability of teaching practices. Despite its critical relevance, comparative research mostly compared teaching practices profiles across vastly different countries, without first demonstrating cross-cultural data comparability (e.g., OECD, 2013). This is an important omission that our study aims to remedy.

Unlike teacher-directed and student-centred practices, classroom assessment has been highlighted more recently in the international debate on teaching practices. Echazarra and colleagues (2016) positioned classroom assessment practices between traditional (teacher-directed) and modern (student-centred) approaches to instruction. Yet, it can be expected that specific assessment practices are infused in both types of practices to varying extents: Some assessment practices may show stronger relations to teacher-directed practices, whereas others might be more closely related to student-centred practices, and vice versa. For instance, Klieme (2020) suggests that formative assessment (including feedback) is more strongly related to teacher-directed instruction than to student-centred teaching. Similarly, effectiveness research often describes a combination of classroom assessment with either teacher-directed or student-centred practices (i.e., to structure or individualize instruction) as the most effective tool to boost student learning (OECD, 2013). Thus, operationalizing classroom assessment as a third, separate dimension of teaching

practices (as for instance practiced by PISA, see OECD, 2014) might not be adequate. Instead, classroom assessment practices are expected to be compatible with both teacher-directed and student-centred ways of teaching. Yet, to date, most studies treated classroom assessment as latent-factor based (e.g., Klieme, 2020). Consequently, when integrating classroom assessment practices into teacher-directed and student-centred teaching activities, an alternative nuanced measurement model may be needed to unfold the intricate interactions between practices. With a wealth of data collected in large-scale educational assessment, such an attempt has not been made so far. The lack of theoretical foundation and empirical research thereby calls for an explorative approach. In this study, we explore how *individual* classroom assessment practices relate to teacher-directed and student-centred teaching practices across countries?

2 Method

2.1 Database and sample

We based our analysis on the 2012 cycle of PISA main study data of students' perceptions of teaching practices in mathematics lessons (see OECD, 2014). To ensure sufficient cultural variations and robustness in findings of different psychometric methods on the structure and metrics of teaching practices, we selected four clusters of countries based on main language families and included three countries/economies in each cluster. The selected country clusters not only differ in language, but also in their affluence level, cultural values of individualism-collectivism, and power distance, which have a bearing on the perceptions and preferences of teaching practices. Our chosen German- and English-speaking countries represent high affluence, high individualism, and low power distance cultures stemming from different pedagogical traditions (the German-speaking countries have highly tracked systems, the English-speaking countries have comprehensive school systems), while the Chinese- and Spanish-speaking

countries represent moderately affluent, collectivistic, and high power distance cultures (the Spanish-speaking countries are infrequently studied in international comparisons to date and they add insight beyond the typical West-East comparisons) (Hofstede, 2001). To rule out method artefacts due to missing values and different sample sizes, a random subsample of 1,000 students with complete responses on the targeted teaching practice items per country/economy were drawn. Therefore, analysis was conducted with 1,000 students for each of three Chinese-speaking (Macao[1], Shanghai, Taipei), English-speaking (Australia, United Kingdom, United States), German-speaking (Austria, Germany, Switzerland), and Spanish-speaking (Chile, Colombia, Mexico) countries/economies, respectively (resulting in N= 12,000).

2.2 Measures

In the 2012 PISA, teaching practices encountered by students in mathematics lessons were measured with 13 items (five items for teacher-directed practices and four items each for student-centred and classroom assessment practices). Students responded on a 4-point Likert-type scale ranging from "*Every lesson*" to "*Never or hardly ever*" (see Table 1).

[1] Since Shanghai, Macao, and Taipei were treated as separate educational systems in PISA 2012, we treat them as "countries" in our study for simplicity, even though they should be referred to as cities/educational systems.

Table 1

Items Measuring Teaching Practices in Mathematics Instruction (PISA 2012)

Practices	Item wording	Response scale
Teacher-directed	The teacher sets clear goals for our learning (T1). The teacher asks me or my classmates to present our thinking or reasoning at some length (T2). The teacher asks questions to check whether we have understood what was taught (T3). At the beginning of a lesson, the teacher presents a short summary of the previous lesson (T4). The teacher tells us what we have to learn (T5).	1= Every lesson 2= Most lessons 3= Some lessons 4= Never or hardly ever
Student-Centred	The teacher gives different work to classmates who have difficulties learning and/or to those who can advance faster (S1). The teacher assigns projects that require at least one week to complete (S2). The teacher has us work in small groups to come up with joint solutions to a problem or task (S3). The teacher asks us to help plan classroom activities or topics (S4).	
Classroom Assessment	The teacher tells me about how well I am doing in my mathematics class (A1). The teacher gives me feedback on my strengths and weaknesses in mathematics (A2). The teacher tells us what is expected of us when we get a test, quiz or assignment (A3). The teacher tells me what I need to do to become better in mathematics (A4).	

2.3 Analysis strategy

These teaching practice items were analysed using both factor analysis (to test Hypotheses 1 and 2 with regard to teacher-directed and student-centred practices) and network analysis (to explore the relations of individual classroom assessment practices with teacher-directed and student-centred practices as formulated in the additional research question). All data and analysis codes are provided in the Open Science Framework.

2.3.1 Hypothesis-testing: Identifying teacher-directed and student-centred teaching practices across countries

To test if teacher-directed and student-centred practices are two distinct factors across countries that reach metric invariance as postulated by Hypothesis 1 and 2, we first tested measurement invariance of a two-factor model comprising teacher-directed practices (five Items) and student-centred practices (four Items) in MGCFA across all 12 countries. Afterwards, we ran a three-factor MGCFA across the 12 countries to entertain the possibility that classroom assessment (measured with four items), next to the two factors in the first model, forms a third factor in the teaching practice framework. The model fit is evaluated by Chi-square tests and the Comparative Fit Index (CFI) (above 0.90 acceptable and above 0.95 excellent), the Root Mean Square Error of Approximation (RMSEA) (below 0.08 acceptable) and the Standardized Root Mean Square Residua (SRMR) (below 0.08 acceptable) (Cheung & Rensvold, 2002; Hu & Bentler, 1999). The acceptance of the more restricted model is based on the changes of CFI and RMSEA values in comparison to the less restricted model. In comparisons involving more than 10 groups, Rutkowski and Sventina (2014) proposed to set the cut point of change of CFI to 0.02 and that of RMSEA to 0.03 from the configural to the metric model, and from the metric to the scalar model the changes of both CFI and RMSEA should be within 0.01. We adhere to these criteria. All factor analyses were performed with the “lavaan” package in R (Rosseel, 2011).

2.3.2 Explorative approach: Integrating classroom assessment into the framework of teaching practices

Next, we performed network analysis using the R-package *qgraph* (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012) to explore the structure and co-occurrence of teaching practices across countries. For each of the 12 countries, we estimated and visualized a partial correlation network (i.e., edges are estimated based on partial correlations between two indicators, controlling for all other indicators in the network, *cor_auto* was applied to create the correlation matrices). This analysis also incorporated a regression-based filtering approach, the "least shrinkage and selection operator" (LASSO), which leads to the estimation of a sparse, more interpretable model (with the hyperparameter gamma set to 0.50 for all models). Consequently, the absence of a connection (i.e., edge) represents conditional independence between two indicators (Christensen et al., 2018). To ensure the accuracy and stability of the estimates, a nonparametric bootstrapping test was performed for each country (i.e., for edges and the centrality index) (Epskamp & Fried, 2017). For these country-specific networks, we conducted three sets of analysis.

Network Comparison Test. First, we performed pair-wise comparisons (= 66 comparisons) of the invariance of the *overall network structure* (operationalized as connection strength matrix) and the *global connectivity* (operationalized as weighted sum of absolute connections). They together inform about the similarity and differences of teaching practices with regard to global patterning across the country-specific networks. This was done with the significant testing based on permutations in the *NCT* package in R (Network Comparison Test, NCT, see Van Borkulo et al., 2016) with the LASSO regularization in which the hyperparameter was set to 0.50.

Edge differences. In each network, we compared the *edge differences* of the individual classroom assessment nodes to teacher-directed and student-centred practices, to clarify if

individual classroom assessment practices are significantly more or less associated with (one of) the two established teaching practices. This was done in the R-package *bootnet* (Epskamp & Fried, 2017), with bootstrapped edge differences plotted out and their significance summarized.

Node centrality. Thirdly, we checked the similarity and differences in the importance of specific nodes in the country-specific networks. We focused on *strength-centrality* (sum of all edge weights connected to a given node in weighted networks). Nodes with higher values of strength-centrality influence other nodes more strongly, without considering other mediating nodes (Costantini et al., 2015). Other centrality indexes such as closeness and betweenness were not targeted, given their lower reliability and reproducibility in comparison to strength-centrality (Fried et al., 2018).

3 Results

3.1 Hypothesis-testing: Identifying teacher-directed and student-centred teaching practices across countries

A MGCFA was performed on the nine items distinguishing teacher-directed and student-centred practices across all 12 countries. The model fit (see Table 2) points to acceptable metric invariance (acceptable CFI, RMSEA, and SRMR values for the metric model and drop of CFI and RMSEA values within the cut-off values of 0.02 and 0.03 respectively), which indicated a universal factor structure of teachers' practices with one factor for teacher-directed and one for student-centred practices, respectively. Hypothesis 1 was supported. In the metric invariance model, the factor loadings for teacher-directed practice items ranged from 0.54 to 0.63, and for the student-centred practice items from 0.55 to 0.64, suggesting that these items were relatively comparable indicators for the two constructs. However, scalar invariance was not supported, which was not unexpected (Hypothesis 2 supported). This may be due to intrinsic differences in metrics

of these constructs across cultures or methodological artefacts that prevented valid cross-country comparisons on mean levels of the two constructs.

Table 2

Model Fit of Measurement Invariance Tests for Teacher-Directed and Student-Centred Practices in Multigroup Confirmatory Factor Analysis

	χ^2	df	CFI	RMSEA	SRMR
Configural	1882.911**	312	.927	.071	.045
Metric	*2178.765***	*389*	*.907*	.069	.070
Scalar	11635.861**	964	.740	.109	.089

Note. Most restrictive model with acceptable fit is printed in italics. ** $p < .01$.

In the three-factor MGCFA model, in which teacher-directed and student-centred practices and classroom assessment were distinguished, only the configural model was just accepted [χ^2 (744, N=12,000) = 5084.600, $p < 0.01$, CFI =0.900, RMSEA =0.076, SRMR =0.050], and the factor loadings, item intercepts, and the associations among the three factors differed enormously across countries, pointing to a lack of support for the comparable three-factor solution across countries.

3.2 Explorative approach: Integrating classroom assessment into the framework of teaching practices

To explore how classroom assessment practices relate to teacher-directed and student-centred practices across countries, partial correlation networks were estimated for each of the 12 countries (see Figure 1a-l). Since the MGCFA for teacher-directed and student-centred practices demonstrated metric invariance, the construct scores (operationalized as the rounded mean scores

across items measuring each construct) were used as nodes in the networks (nodes TD and SC) together with the classroom assessment items (Nodes A1-A4). These two construct scores can be considered ordered categories with the same metric as the classroom assessment items. The nonparametric bootstrapping testing based on 1000 bootstrapped samples showed support for the accuracy of the networks. The strength centrality indexes also showed acceptable stability, with the stability coefficients (CS cor =0.70) all over 0.50 except for the US and Chile (both still over 0.25, with a value of 0.44 and 0.36, respectively). Supplement 1 presents all graphs for the recovery of the edges per country and Supplement 2 presents a table of the correlation stability coefficient [CS(cor=0.70)] per country.

3.3 Overall network structure and global connectivity

A visual inspection of the 12 country-specific networks revealed that most edges were positive, indicating that the more frequent application of one practice seems to go hand in hand with the more frequent application of another connected practice, conditioning on all remaining practices. Even the TD and SC nodes were positively connected across countries, with relatively stronger edge weights in the networks for the Spanish- and German-speaking countries (weights between 0.19 and 0.27), and comparably weaker edge weights in the networks for the Chinese-speaking countries (weight Macao=0.15 and Taipei=0.17), and particularly in Shanghai (weight=0.09). A few exceptions of negative edges (dashed lines) were observed in the networks for Taipei (SC and A4), the US and Austria (both SC and A3), and Chile and Mexico (both TD and A2), and all these edges were weak (weights between -0.06 and -0.10).

Figure 1

Country-specific Partial-Correlation Networks of Teaching Practices

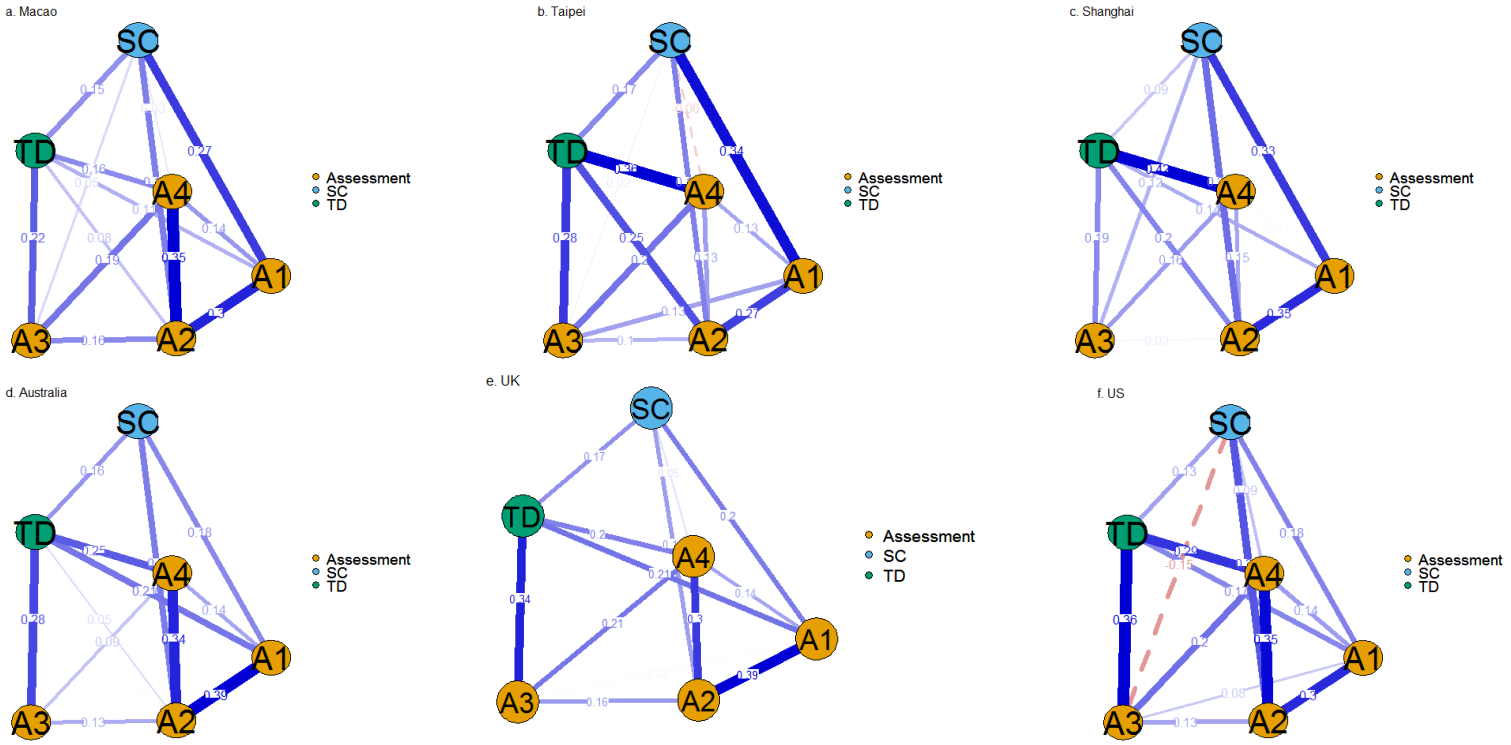
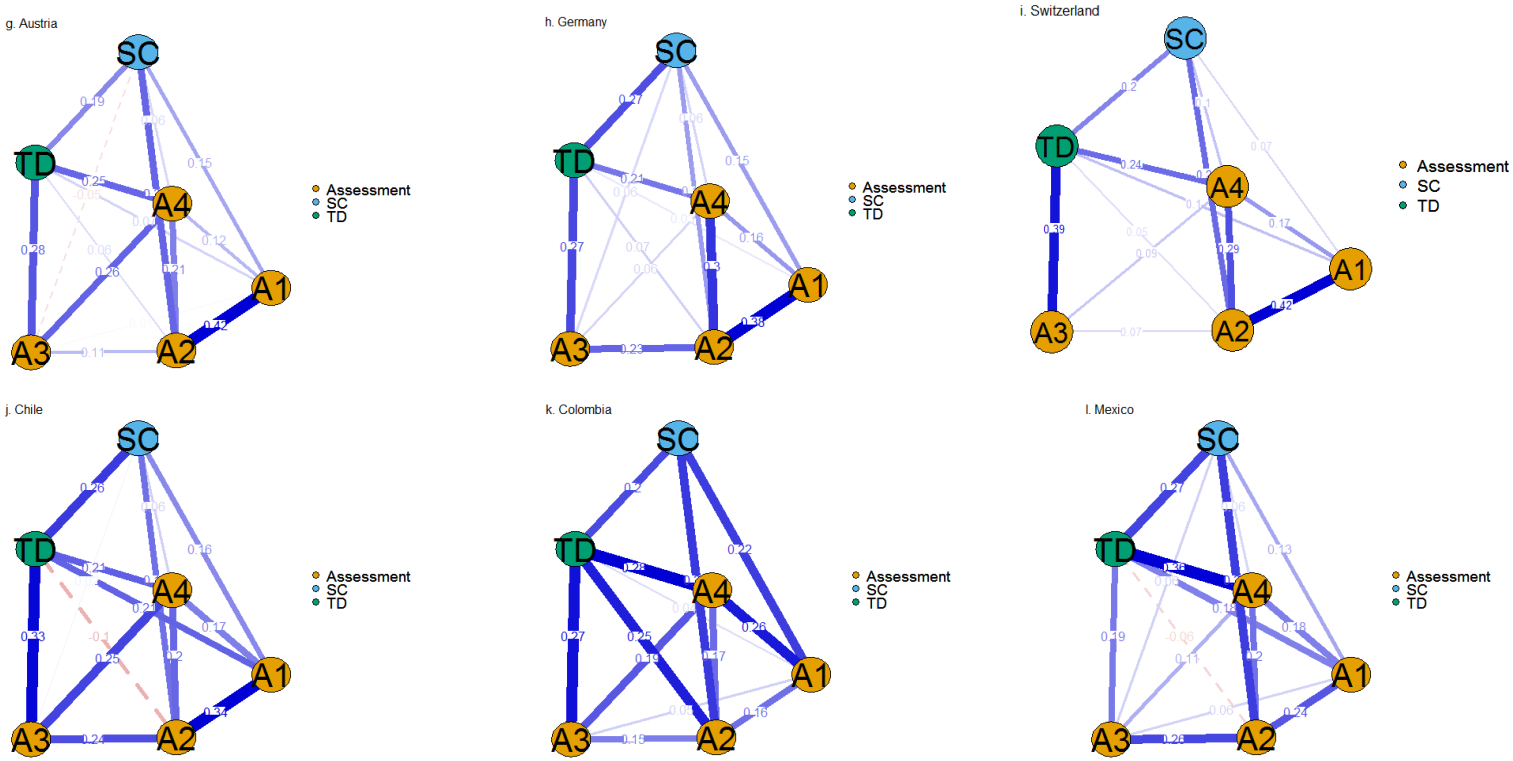


Figure 1 (continued)

Country-specific Partial-Correlation Networks of Teaching Practices



Note. Partial correlation networks of teaching practices with rounded mean-scores for teacher-directed (TD) and student-centred (SC) practices, individual items for classroom assessment. Full (blue) lines represent positive edges; dashed (red lines) represent negative edges. To facilitate visualization, the position of the nodes is the same across networks (Germany is reference country). A1= feedback performance in class, A2= feedback individual strength and weaknesses, A3= informing about expectations in test, A4= feedback how to improve.

NCT: Network structure. Table 3 presents the results of the pairs-wise tests of the overall network structure invariance (M-test statistics above the diagonal) and global connectivity invariance (S-test statistics below the diagonal). The overall network structure significantly differed for 48 out of the 66 pair-wise comparisons ($p < 0.05$). Among countries belonging to the same linguistic/cultural cluster, the overall network structure invariance (i.e., comparability) was supported for all three English-speaking countries, Switzerland and Austria, Shanghai and Taipei, and Mexico and Chile. For countries belonging to different linguistic/cultural clusters, the network structures mostly differed. Exceptions were for Switzerland and all three English-speaking countries; Austria and the US and UK; and Macao and Chile and two German-speaking countries each (Macao: Austria and Germany; Chile: Austria and Switzerland). The network structure for Germany, Shanghai, Mexico, and Colombia showed the least comparability, with only one invariant comparison each. Even though not supported for all within-cluster pair-wise comparisons, it seems that there was a more similar network structure for countries belonging to the same linguistic/cultural cluster (especially the English-speaking cluster) than of countries belonging to different linguistic/cultural clusters.

NCT: Global connectivity. With regard to the pair-wise comparisons of the global connectivity, the US network was significantly more connected (i.e., these teaching practices frequently co-occurred) than the networks for most other countries. The same applied to the network for Chile (compared to Austria, Shanghai, and Macao), and Mexico (compared to Germany, Shanghai, and Macao). The US network showed the comparably highest global strength ($S=2.55$), followed by Mexico ($S=2.39$), and Chile ($S=2.37$), whereas the networks for the Chinese- and German-speaking countries showed a comparably low global connectivity (for global strength indices per country see Table 4).

Table 3

Results Pair-wise Network Comparison Test (NCT) for Network Structure and Global Connectivity Invariance

	AUT	GER	CHE	AUS	UK	US	CNQ	TAP	MAC	COL	CHL	MEX
AUT		0.15**	0.12	0.13**	0.11	0.11	0.20**	0.15**	0.12	0.21**	0.10	0.17**
GER	0.06		0.15**	0.12**	0.13**	0.15**	0.19**	0.15**	0.10	0.17**	0.15**	0.14**
CHE	0.10	0.03		0.11	0.09	0.08	0.23**	0.19**	0.14**	0.19**	0.13	0.16**
AUS	0.09	0.03	0.01		0.10	0.09	0.16**	0.17**	0.10	0.18**	0.14**	0.15**
UK	0.11**	0.05	0.02	0.02		0.09	0.18**	0.19**	0.10**	0.19**	0.10	0.15**
US	0.34**	0.28**	0.25**	0.25**	0.23**		0.19**	0.19**	0.14**	0.19**	0.15**	0.17**
CNQ	0.07	0.01	0.03	0.02	0.04	0.27**		0.10	0.21**	0.20**	0.20**	0.20**
TAP	0.12	0.06	0.02	0.03	0.01	0.22**	0.05		0.16**	0.13	0.22**	0.22**
MAC	0.05	0.02	0.05	0.04	0.07	0.30**	0.02	0.07		0.15**	0.14**	0.15**
COL	0.11	0.05	0.02	0.02	0.00	0.23	0.04	0.01	0.07		0.22**	0.22**
CHL	0.16**	0.09**	0.06	0.07	0.04	0.19**	0.09	0.03	0.11**	0.04		0.12
MEX	0.18**	0.12**	0.08	0.09	0.07	0.16**	0.11**	0.06	0.13**	0.07	0.02	

Note. Above diagonal: network structure invariance (M-statistic), below diagonal: network global connectivity invariance (S-statistic); **= significantly different network structure, global connectivity if $p < 0.05$; AUS= Australia, AUT=Austria, CHE=Switzerland, CHL=Chile, COL=Colombia, GER=Germany, UK= Great Britain, MAC= Macao, MEX=Mexico, QCN= Shanghai, TAP= Taipei, US= United States.

To summarize, these country-specific networks not only mostly significantly differed with regard to network structure (to a lesser extent for countries belonging to the same linguistic/cultural cluster), but also in global connectivity.

Table 4

Global Network Connectivity per Country

Country	Global strength
Macao	2.25
Taipei	2.33
Shanghai	2.28
Australia	2.30
UK	2.32
US	2.55
Austria	2.21
Germany	2.27
Switzerland	2.31
Chile	2.37
Colombia	2.32
Mexico	2.39

3.4 Relation of classroom assessment practices to TD and SC

We paid special attention to the individual classroom assessment nodes and their relations (i.e., edges) to the TD and SC nodes per country. A visual inspection suggested that the classroom assessment nodes did not cluster strongly, but showed rather distinct partial correlations with either TD or SC. For each country-specific network, bootstrapping was performed to test if individual classroom assessment practices significantly differently related to either TD or SC. The bootstrapped differences between all edge weights were plotted out and are presented in

Supplement 3. We summarized the significance of edge differences between each of the four classroom assessment nodes and TD and SC in Table 5.

In all country-specific networks (except for Macao) A4 (informing individual students about what is needed to become better in mathematics) was more strongly, conditionally related to TD than SC. Similarly, A3 (informing what is expected of the class in tests or assignments) exhibited a significantly stronger unique relation with TD compared to SC in all countries, except in Shanghai and Mexico. The remaining two classroom assessment practices (A1: informing about the performance in mathematics class; A2: giving individual feedback on strength and weaknesses) showed some ambiguous relations to TD and SC. In all Chinese-speaking countries (as well as in Columbia), A1 was significantly more strongly, conditionally related to SC than TD, whereas in all other countries, these two edge differences were not significant. Among the non-Chinese speaking countries, A2 was more strongly conditionally linked with SC than TD (two other exceptions were Germany and Columbia, where no significant difference between the edge weights was observed). Thus, the four classroom assessment nodes did not cluster together strongly; they rather exhibited different relations to either teacher-directed or student-centred teaching practices, as detailed above.

3.5 Strength-centrality of individual nodes

In a next step, we investigated the strength-centrality of individual nodes within each country-specific network (see Figure 2). Across countries, informing on individual strength and weaknesses in mathematics (A2) seemed to play a central role (average strength: 1.03) followed by the teacher-directed node (average strength: 0.96), and the two assessment practices telling individual students what is needed to become better in mathematics (A4, average strength = 0.88), and informing students about how they are performing in their mathematics class (A1, average

strength = 0.82). The student-centred and A3 node (what the class needs for a test, quiz, or assignment) played a less central role across countries (average strength: 0.67 and 0.69, respectively). The remaining nodes varied with regard to their importance across countries (particularly A3 with strength centrality values between 0.50 and 0.92). The country-specific strength-centrality of individual nodes can be found in Supplement 4.

Table 5

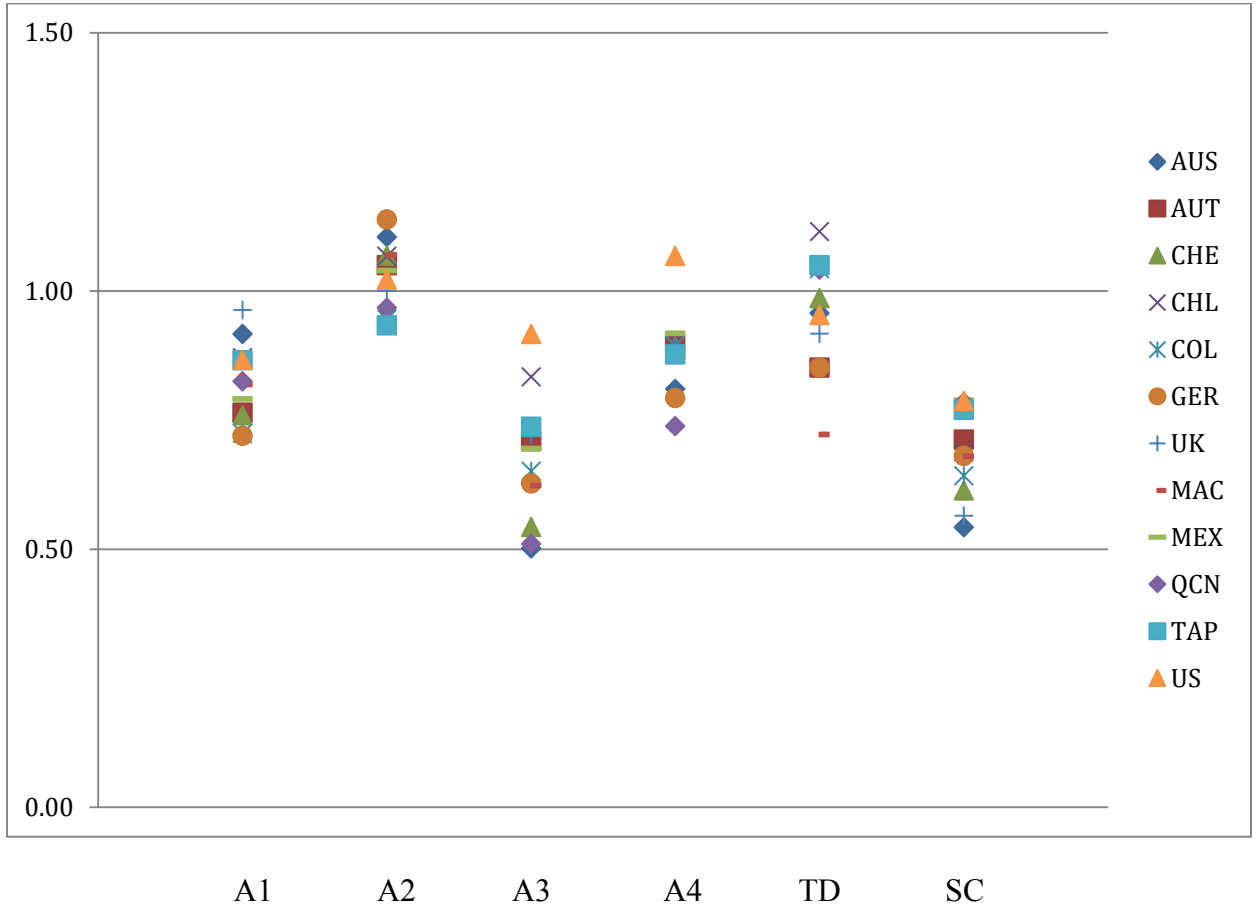
Significance of Edge Difference Tests of Each Classroom Assessment Node (A) with Teacher-Directed (TD) and Student-Centred (SC) Nodes

Edges compared	AUT	GER	CHE	AUS	UK	US	CNQ	TAP	MAC	COL	CHL	MEX
A1-TD vs A1-SC	X	X	X	X	X	X	V	V	V	V	X	X
A2-TD vs A2-SC	V	X	V	V	V	V	X	X	X	X	V	V
A3-TD vs A3-SC	V	V	V	V	V	V	X	V	V	V	V	X
A4-TD vs A4-SC	V	V	V	V	V	V	V	V	X	V	V	V

Note. V indicates significant edge difference at $p < 0.05$; X indicates nonsignificant edge difference at $p < 0.05$. A1= feedback performance in class, A2= feedback individual strength and weaknesses, A3= informing about expectations in test, A4= feedback how to improve, TD= rounded mean score for teacher-directed practices, SC=rounded mean score for student-centred practices. AUS= Australia, AUT=Austria, CHE=Switzerland, CHL=Chile, COL=Colombia, GER=Germany, UK= Great Britain, MAC= Macao, MEX=Mexico, QCN= Shanghai, TAP= Taipei, US= United States.

Figure 2

Strength Index of the Partial-Correlation Networks across Countries



Note. AUS= Australia, AUT=Austria, CHE=Switzerland, CHL=Chile, COL=Colombia, GER=Germany, UK= Great Britain, MAC= Macao, MEX=Mexico, QCN= Shanghai, TAP= Taipei, US= United States. A1= feedback performance in class, A2= feedback individual strength and weaknesses, A3= informing about expectations in test, A4= feedback how to improve, TD= rounded mean score for teacher-directed practices, SC=rounded mean score for student-centred practices.

4 Discussion

We set out to investigate the cross-cultural similarities and differences in the structure and co-occurence of teaching practices in mathematics instruction with a 12-country dataset from a large-scale international survey (PISA). We combined factor analysis and network analysis to test our hypothesis and research question. Rooted in instructionist and constructivist theories of teaching (Tobias & Duffy, 2010), the distinction between teacher-directed and student-centred teaching practices and their similar structure but not origin of metrics (i.e., item intercepts) across cultures were postulated (Hypothesis 1 and 2). Given the lack of theory and empirical foundation, we additionally explored how classroom assessment practices position within the broad range of teaching practices and investigated how individual assessment practices differently related to either teacher-directed or student-centred teaching practices.

We confirmed metric but not scalar invariance of teacher-directed and student-centred practices in the MGCFA of students' self-reported frequency of practices across countries (supporting Hypothesis 1 and 2). Adding classroom assessment as a third factor in the MGCFA did not support an invariant three-factor structure across countries; whereas a network analysis per country on individual classroom assessment practices and the rounded mean scores of teacher-directed and student-centred practices showed rather different direct interactions among the teaching practices. Network analyses revealed that (1) across countries, most teaching practices were positively mutually linked (even teacher-directed and student-centred practices), (2) the overall network structure and to a lesser extent global connectivity differed for most pair-wise comparisons, but similarity of the network structure was often found for countries belonging to the same cultural and linguistic cluster; (3) the classroom assessment items did not form a cluster and do not seem to be latent-factor based and among the four classroom assessment practices, A4

(informing individual students about what is needed to become better in mathematics) and A3 (informing what is expected of the class in tests or assignments) more strongly related to teacher-directed practices than student-centred practices across countries, whereas the other two classroom assessment practices (A1: informing on how the student is doing in the mathematics class; A2: feedback on individual strength and weaknesses) showed less common patterning in their relation to either teacher-directed or student-centred practices, but tended to be more mutually linked to student-centred practices, and (4) in comparisons of the relative importance of specific practices in the country-specific networks, A2 (informing on individual strength and weaknesses in mathematics) and the node for teacher-directed practices played a relative important role on average across countries, whereas the node for student-centred practices was less important. In the following, we discuss the global patterning and implications. We refrain from diving into specifics of cross-country differences, given that no clear expectation was formulated and the exploratory nature of the analysis.

4.1 Teacher-directed and student-centred practices: Two distinct approaches to teaching?

Theoretically, teacher-directed and student-centred practices are based on two distinct - and even often labelled as opposite - approaches to instruction. Our MGCFA supports this theoretical distinction across countries. Thus, the theories of instruction developed and tested in Western countries are generalizable to the non-Western countries in our study (e.g., East-Asian and Latin-American countries). However, the consistently positive conditional relation between teacher-directed and student-centred practices in our country-specific networks highlights that teachers do not stick to only one approach to teaching, but combine practices stemming from different teaching traditions. Even within a lesson students are likely to be exposed to various teaching practices (Echazarra et al., 2016). Thus, teacher-directed and student-centred practices

complement each other to fit the context, subject content, and students. Consequently, the strict theoretical distinction might not reflect the more flexible co-occurrence of teaching practices in reality. This seems to be less the case for the Chinese-speaking countries, where we observed a comparably low, yet positive relation between teacher-directed and student-centred practices (i.e., less frequent co-occurrence of teacher-directed and student-centred practices). One possible explanation is that East-Asian countries value conformity and legitimize power distance more than the other linguistic/cultural clusters of countries (Hofstede, 2001), thus they tend to strictly adhere to one specific instructional approach (i.e., traditional teacher-directed instruction, see Echazarra et al., 2016).

4.2 Integrating classroom assessment into the framework of teaching practices

Our network analysis on the structure and co-occurrence of teaching practices challenges the proposal to conceptualize classroom assessment practices as a third set of practices as well as the positioning between traditional (teacher-directed) and modern (student-centred) approaches to teaching (Echazarra et al., 2016). This characterisation might be an oversimplification of the complex nature of classroom assessment. Instead of clustering together, these individual classroom assessment practices tended to show a stronger relation to either teacher-directed or student-centred practices. A more teacher-directed approach to instruction is clearly related to assessment practices that are used to structure and guide classroom learning, such as informing students about learning goals (i.e., what is expected in tests, a quiz, or assignments) or providing advice on how to reach specific goals (i.e., what is needed to become better in mathematics). A more student-centred approach to instruction, on the other hand, tends to be related to assessment practices supporting individualized learning, such as providing individual feedback on strength and weaknesses or feedback with a social reference frame (i.e., how well a student is doing in

mathematics class). Thus, network analysis provides a more nuanced look on the relation between individual assessment practices and teacher-directed or student centred-practices. It should be noted that both directions of the relation are possible: i.e., the specific approaches to teaching lead to the co-occurrence of specific assessment practices or vice versa. Consequently, it is plausible that these classroom assessment practices do not stem from one tradition, but are instilled in teaching from multiple traditions. Moreover, cross-cultural differences on strength of the links add complexity to the picture. In any case, treating them as one factor would obscure these nuanced differences. Moreover, our results emphasize the broad nature of the concept *teaching practices*, intertwining practices stemming from multiple teaching traditions with a complex relation to each other. We urge further research to define the theoretical concept more precisely.

4.3 The structure and dynamics of teaching practices across countries

Across countries, we found mostly different network structures and global connectivity. However, we also found an invariant structure of the networks among the three English-speaking countries; Taipei and Shanghai; Austria and Switzerland; and Mexico and Chile, indicating more similarity within linguistic/cultural clusters of countries than across clusters (this is in line with findings of Fischer, Praetorius, & Klieme, 2019). The cultural and colonial heritage of the three English speaking countries, their shared teaching traditions and (comprehensive) school system structure seem to be more similar than in the other linguistic/cultural clusters, which may contribute to higher levels of similarity of the networks (the comparability of teaching constructs for English-speaking countries was also demonstrated in other studies, see Fischer et al., 2019 or Klieme, 2020). Across linguistic/cultural clusters, interestingly, Switzerland's network structure was comparable to the structure of all three English-speaking countries, and also Austria's network showed an invariant structure compared to the UK and US. Thus, German- and English- speaking

countries might be relatively similar in terms of teaching culture, compared to Chinese- and Spanish-speaking countries. To draw valid conclusions, future research should investigate similarities and differences between countries in more detail (e.g., with regard to school systems, preferred teaching approaches, but also cultural and colonial traditions). Moreover, countries also differ with regard to the importance of different teaching and assessment practices and the relation of individual assessment practices with either teacher-directed or student-centred practices in particular (see previous section). Our results emphasize context-specific structures and patterns of teaching practices. Consequently, targeted interventions have to be tailored to the specific context in order to be effective in the respective countries and should not be overly generalized or "borrowed" across countries.

4.4 Centrality of teaching practices: Starting point for targeted interventions

With strength centrality indices, we also witness the relative importance of individual practices within a network of teaching practices. We view the most central practices as the binding practices in the teaching practice networks. In other words, they are versatile because they can accompany many other practices and are easily aligned with other practices. It is our extrapolation that increasing the central practice may facilitate promoting other practices to be used in combination. We thus expect that a stimulation of the most central practice is beneficial as it might influence many other practices that are well linked with it. Classroom assessment practices with a focus on individual students - particularly the practice of providing individualized feedback on strength and weaknesses (A2, most important node on average across countries) and the practice of giving individual advice on how to get better (A4) seem to be at the heart of teaching as perceived by students. In contrast, assessment practices focusing on the class (A3: the teacher tells us what is expected in a quiz or assignment, A1: the teacher compares me with my class) seem to

be less influential on other teaching practices. Similarly to findings of Echazarra and colleagues (2016), teacher-directed practices seem to be more important in mathematics instruction than student-centred practices in our study.

4.5 Toolbox for measurement investigations

Methodologically, empirically testing measurement invariance of constructs in large-scale surveys before drawing any cross-cultural comparison is important in order to ensure the level of comparability and draw valid comparative inferences (Boer, Hanke, & He, 2018). Psychometric tools abound (e.g., item response theory-based scaling, latent class analysis), and flexible applications are in much need. We made use of two methods for different purposes. MGCFA was used for theory testing and confirmation, whereas network analysis was resorted to aid measurement in exploratory ways. MGCFA with its various adaptations and extensions (e.g., partial invariance, approximate invariance testing, or simultaneous mixture CFA) provides rigorous and realistic testing of measurement of multiple-item measures. Network analysis is especially useful for relatively new constructs with less clear conceptualizations (e.g., classroom assessment practices) and that may not be latent-factor based (Costantini et al., 2015). These tools complement each other and deepen our understanding on substantive educational phenomena, as they either capture the commonality (MGCFA) or the unique interactions not accounted by the commonality (network analysis). For network analysis, there is a new development towards a better integration with classic psychometrics (Epskamp et al., 2017), and new research questions can be answered with information gathered in network analysis (e.g., what combination or dynamics of teaching practices especially contribute to student learning, how global connectivity in partial correlation networks of teaching practices is related to national policy on teacher autonomy).

4.6 Limitations

When interpreting the results of our study, some limitations have to be considered. Firstly, we used PISA data, where students are nested within schools (without clustering at classroom level). Self-reports of students taught in possibly different classrooms by different teachers ignore the heterogeneity at classroom levels, and thus have inferential limits. This is unfortunate as the interpretation of many aspects of instruction is not only located on the individual but also on the class level (Lüdtke, Robitzsch, Trautwein, & Kunter, 2009). Future research should use multiple data sources (especially teachers' self-report and observations in real classes) to validate our results. Secondly, potential measurement bias in item responses (e.g., translation errors, misinterpretation of item content) may be detected in MGCFA, but still may exist in network analysis, which can challenge the validity of comparisons of structure, edge weights, and centrality indices across countries. Other psychometric tools and qualitative procedures are in need to further uncover bias that can limit data comparability. Thirdly, to facilitate comparisons, we randomly selected 1000 students per country. Replications with different subsamples per country or additional country clusters may be performed to check the robustness of our results. And lastly, following the results of the MGCFA (identifying two separate factors across countries) - we included teacher-directed and student-centred teaching practices as rounded mean construct scores in our network analysis. Thus, we had no information on which specific teacher-directed and student-centred practices are interlinked and the strength of their connection. Further research should investigate under which circumstances teachers combine which teaching practices as well as the effectiveness.

5 Conclusion

We have made use of data of representative student samples from multiple countries and complementary psychometric methods to study the structure and co-occurrence of teaching

practices from a cross-cultural perspective. Our empirical support for the distinction between teacher-directed and student-centred practices, and the nuanced differences in classroom assessment practices related to these two well-established teaching practices open up for new perspectives to conceptualize dimensions of teaching practices. We urge researchers to apply innovative measurement models, and expand the measurement to other facets beyond the quantitative focus.

Supplementary data

All data and syntax used in this study are available at https://osf.io/e4fx6/.

References

Abacioglu, C. S., Isvoranu, A.-M., Verkuyten, M., Thijs, J., & Epskamp, S. (2019). Exploring multicultural classroom dynamics: A network analysis. *Journal of School Psychology, 74.* doi:https://doi.org/10.1016/j.jsp.2019.02.003

Astin, W. A., Banta, W. T., Cross, P. K., El-Khawas, E., Ewell, T. P., Hutchings, P., . . . Wright, D. B. (1996). *Nine principles of good practice for assessing student learning*. American Association for Higher Education (AAHE).

Boer, D., Hanke, K., & He, J. (2018). On detecting systematic measurement error in crosscultural research: A review and critical reflection on equivalence and invariance tests. *Journal of Cross-Cultural Psychology,* 49, 713–734. https://doi.org/10.1177/0022022117749042

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation & Accountability, 21,* 5–31. doi:https://doi.org/10.1007/s11092-008-9068-5

Caro, D. H., Lenkeit, J., & Kyriakides, L. (2016). Teaching strategies and differential effectiveness across learning contexts: Evidence from PISA 2012. *Studies in Educational Evaluation, 49,* 30–41. doi:https://doi.org/10.1016/j.stueduc.2016.03.005

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64,* 723–733.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255. doi:https://doi.org/10.1207/s15328007sem0902_5

Christensen, A. P., Kenett, Y. N., Aste, T., Silvia, P. J., & Kwapil, T. R. (2018). Network structure of the Wisconsin Schizotypy Scales-Short Forms: Examining psychometric network filtering approaches. *Behavior Research Methods, 50*, 2531-2550 doi:https://doi.org/10.3758/s13428-018-1032-9

Coombs, A., DeLuca, C., LaPointe-McEwan, D., & Chalas, A. (2018). Changing approaches to classroom assessment: An empirical study across teacher career stages. *Teaching and Teacher Education, 71*, 134-144. doi:https://doi.org/10.1016/j.tate.2017.12.010

Costantini, G., Epskamp, S., Borsboom, D., Perugini, M., Mõttus, R., Waldorp, L. J., & Cramer, O. J. (2015). State of the aRt personality research: A tutorial on network analysis of personality data in R. *Journal of Research in Personality, 54,* 13–29. doi:https://doi.org/10.1016/j.jrp.2014.07.003

Creemers, B. P. M., & Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement, 17*, 347-366. doi:https://doi.org/10.1080/09243450600697242

Dalege, J., Borsboom, D., van Harreveld, F., & Maas, H. (2018). A network perspective on attitude strength: Testing the connectivity hypothesis. *Social Psychological and Personality Science*. doi:https://doi.org/10.1177/1948550618781062

Dewey, J. (1929). *My pedagogic creed.* Washington, DC: Progressive Education Association.

Duncker, K., & Lees, L. S. (1945). On problem-solving. *Psychological Monographs,58,* i-113. doi:https://doi.org/10.1037/h0093599

Echazarra, A., Salinas, D., Méndez, I., Denis, V., & Rech, G. (2016). How teachers teach and students learn: Successful strategies for school. *OECD Education Working Paper*. Paris: OECD Publishing.

Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods.* doi:https://doi.org/10.1037/met0000167

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika, 82*, 904–927. doi:https://doi.org/10.1007/s11336-017-9557-x

Epskamp, S., & Fried, E. I. (2017). bootnet: *Bootstrap methods for various network estimation routines.* Retrieved from https://cran.r-project.org/web/packages/bootnet/index.html

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software, 48,* 1–18. doi:http://dx.doi.org/10.18637/jss.v048.i04

Fischer, J., Praetorius, A.-K., & Klieme, E. (2019). The impact of linguistic similarity on cross-cultural comparability of students' perceptions of teaching quality. *Educational Assessment Evaluation and Accountability,* 31, 201–220. https://doi.org/10.1007/s11092-019-09295-7

Fried, E. I., Eidhof, M. B., Palic, S., Costantini, G., Huisman-van Dijk, H. M., Bockting, C. L. H., . . . Karstoft, K.-I. (2018). Replicability and generalizability of Posttraumatic Stress Disorder (PTSD) networks: A cross-cultural multisite study of PTSD symptoms in four

trauma patient samples. *Clinical Psychological Science, 6*, 335–351. doi:https://doi.org/10.1177/2167702617745092

Fuller, B., & Clarke, P. (1994). Raising school effects while ignoring culture? Local conditions and the influence of classroom tools, rules, and pedagogy. *Review of Educational Research, 64,* 119–157. doi:https://doi.org/10.3102/00346543064001119

Harlen, W. (2007). Formative classroom assessment in science and mathematics. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 116–135). New York/London: Teachers College Press, Columbia University.

Harlen, W., & Deakin-Crick, R. (2002). *A systematic review of the impact of summative assessment and tests on students' motivation for learning*. London: EPPI-Centre.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Hofstede, G. (2001). *Culture's consequences: comparing values, behaviors, institutions, and organizations across nations*. Thousand Oaks, CA: Sage.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6,* 1-55. doi:https://doi.org/10.1080/10705519909540118

Isikoglu, N., Basturk, R., & Karaca, F. (2009). Assessing in-service teachers' instructional beliefs about student-centered education: A Turkish perspective. *Teaching and Teacher Education, 25,* 350–356. doi:https://doi.org/10.1016/j.tate.2008.08.004

Jacobs, J. K., Hollingsworth, H., & Givvin, K. B. (2007). Video-based research made "easy": Methodological lessons learned from the TIMSS video studies. *Field Methods, 19*, 284-299. doi:https://doi.org/10.1177/1525822X07302106

Kippers, W. B., Wolterinck, C. H. D., Schildkamp, K., Poortman, C. L., & Visscher, A. J. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and Teacher Education, 75*, 199-213. doi:https://doi.org/10.1016/j.tate.2018.06.015

Klieme, E. (2020). Policies and practices of assessment: A showcase for the use (and misuse) of international large scale assessments in educational effectiveness Research. In J. Hall, P. Sammons, & A. Lindorff (Eds.). *International Perspectives in Educational Effectiveness Research.* Springer.

Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: how to use student ratings of classroom or school characteristics in multilevel modelling. *Contemporary Educational Psychology, 34,* 120–131.

McCormick, J., & Alavi, S. B. (2004). A cross-cultural analysis of the effectiveness of the Learning Organization model in school contexts. *International Journal of Educational Management, 18,* 408–416. doi:https://doi.org/10.1108/09513540410563112

Mostafa, T., Echazarra, A., & Guillo, H. (2018). The science of teaching science: An exploration of science teaching practices in PISA 2015. *OECD Education Working Paper.* Paris: OECD Publishing.

OECD (2013). *Teaching and Learning International Survey TALIS 2013:Conceptual Framework*. Paris: OECD Publishing.

OECD (2014). *PISA 2012 technical report.* Paris: OECD Publishing.

OECD (2015). *PISA 2015 assessment and analytical framework*. Paris: OECD Publishing.

OECD (2019). *Teaching and Learning International Survey TALIS 2018: Technical report.* Paris: OECD Publishing.

Ormrod, J. E. (2012). *Essentials of educational psychology: Big ideas to guide effective teaching* (3rd ed.). Boston: Pearson.

Piaget, J. (1952). *The origins of intelligence in children*. New York, NY: International Universities Press.

Praetorius, A.-K., Klieme, E., Bell, C. A., Qi, Y., Witherspoon, W., & Opfer, D. (2018). *Country conceptualizations of teaching quality in TALIS video: Identifying similarities and differences.* Paper presentation at the annual meeting of the American Educational Research Association, New York.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science, 28,* 4–13. doi:https://doi.org/10.1002/bs.3830280103

Richardson, V. (2003). Constructivist pedagogy. *Teachers College Record, 105,* 1623–1640. doi:https://doi.org/10.1046/j.1467-9620.2003.00303.x

Rosenshine, B. (1976). Classroom instruction. In N. L. Gage (Ed.), *The psychology of teaching methods* (75th ed., pp. 335–371). Chicago, IL: University of Chicago Press.

Rosseel, Y. (2011). Lavaan: An R package for structural equation modeling. *Journal of statistical software 48*. doi:https://doi.org/10.18637/jss.v048.i02

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement, 74*, 31-57. doi: https://doi.org/10.1177/0013164413498257

Sachisthal, M. S. M., Jansen, B. R. J., Peetsma, T. T. D., Dalege, J., van der Maas, H. L. J., & Raijmakers, M. E. J. (2019). Introducing a science interest network model to reveal country differences. *Journal of Educational Psychology, 111*, 1063–1080. doi:https://doi.org/10.1037/edu0000327

Stigler, J., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York, NY: The Free Press.

Tobias, S., & Duffy, T. M. (Eds.). (2010). *Constructivist instruction: Success or failure?* (Digital print). New York, London: Routledge Taylor & Francis Group.

Van Borkulo, C. D., Epskamp, S., & Milner, A. (2016). *NetworkComparisonTest*. https://cran.rproject.org/web/packages/NetworkComparisonTest/NetworkComparisonTest.pdf

Van de Vijver, F. J.R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research.* Thousand Oaks: Sage.

Van de Vijver, F. J. R. (2018). *Talk at the OECD-GESIS seminar: translating and adapting instruments in large-scale assessments*. Paris.

Vieluf, S., Kaplan, D., Klieme, E., & Bayer, S. (2012). *Teaching practices and pedagogical innovations. Evidence from TALIS.* Paris: OECD Publishing.

Vieluf, S., Kunter, M., & van de Vijver, F. J. R. (2013). Teacher self-efficacy in cross-national perspective. *Teaching and Teacher Education, 35*, 92-103. doi:10.1016/j.tate.2013.05.006

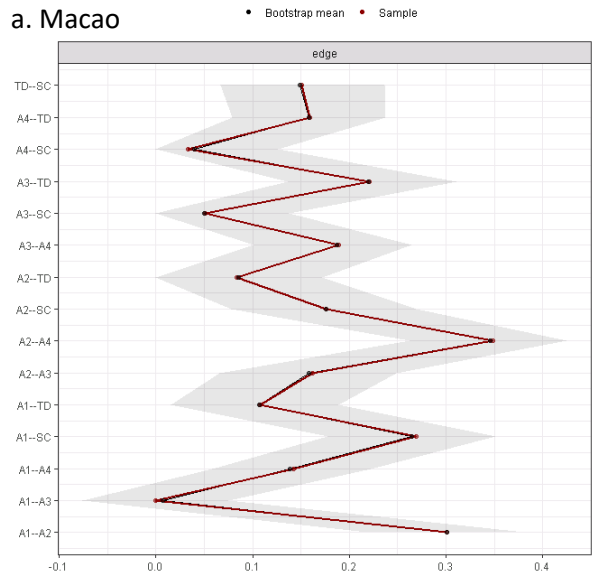
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Zurita, G., & Nussbaum, M. (2004). A constructivist mobile learning environment supported by a wireless handheld network. *Journal of Computer Assisted Learning, 20,* 235–243. doi: https://doi.org/10.1111/j.1365-2729.2004.00089.x

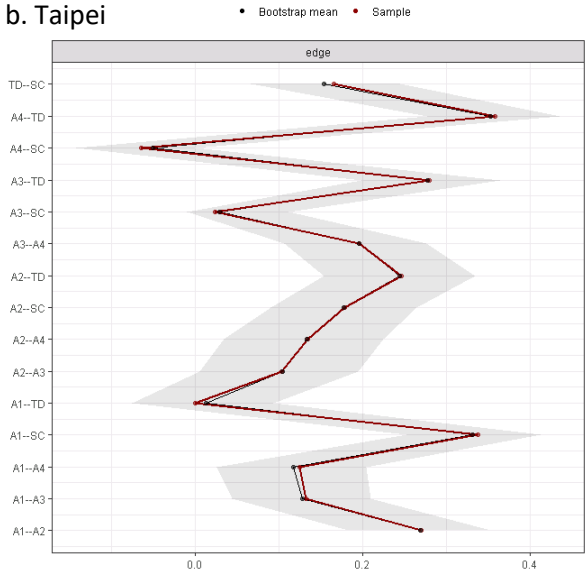
Supplement 1

Edge Recovery across countries - Bootstrapping

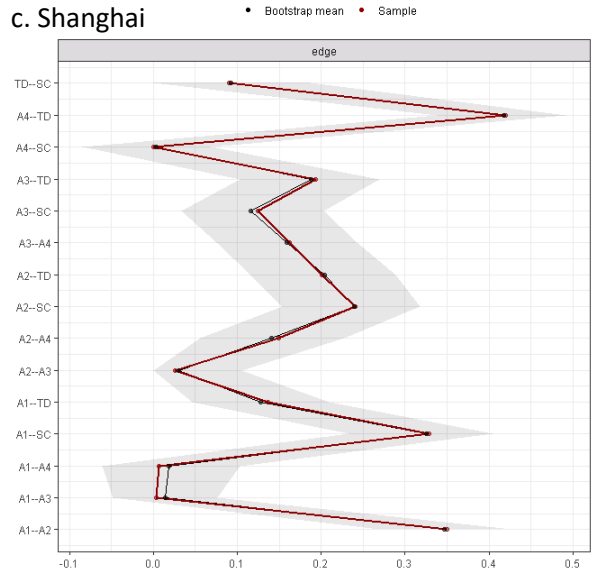
a. Macao



b. Taipei



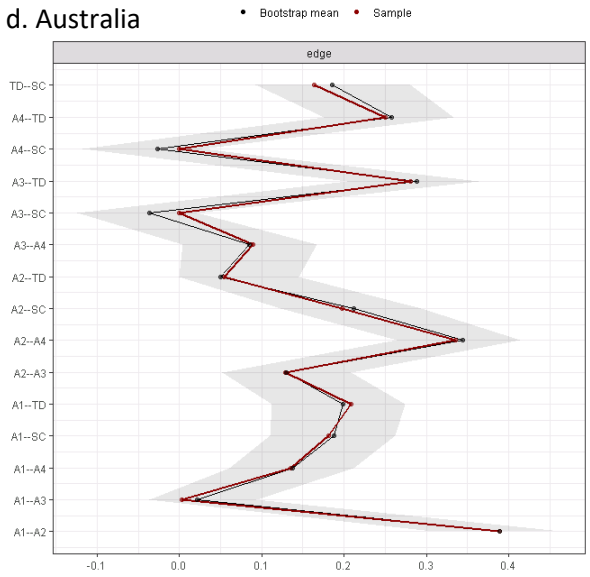
c. Shanghai



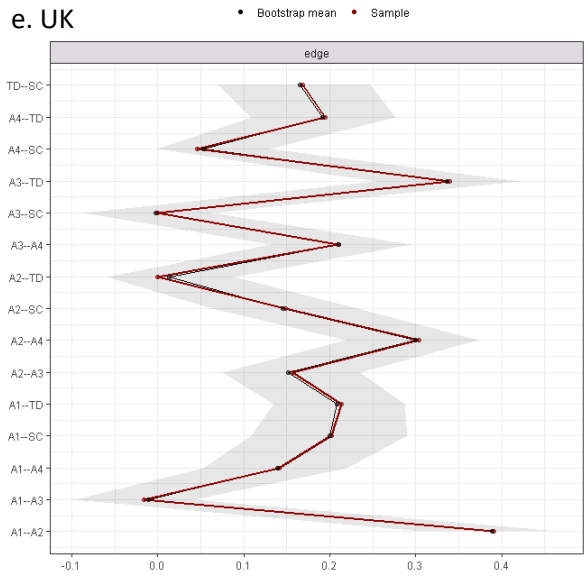
Supplement 1 (continued)

Edge Recovery across countries - Bootstrapping

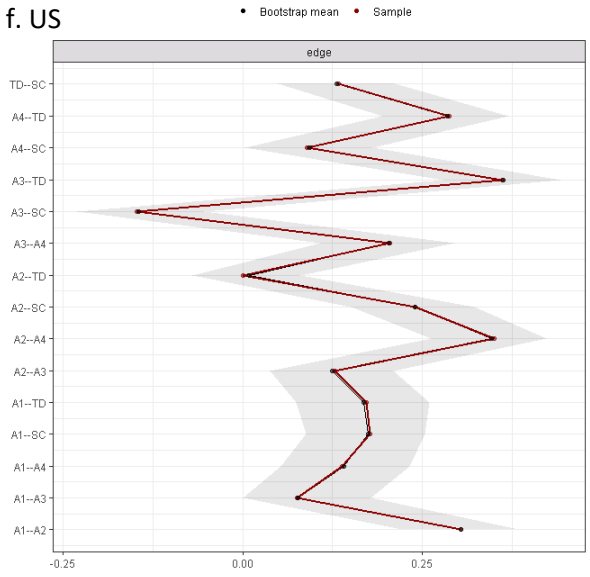
d. Australia



e. UK



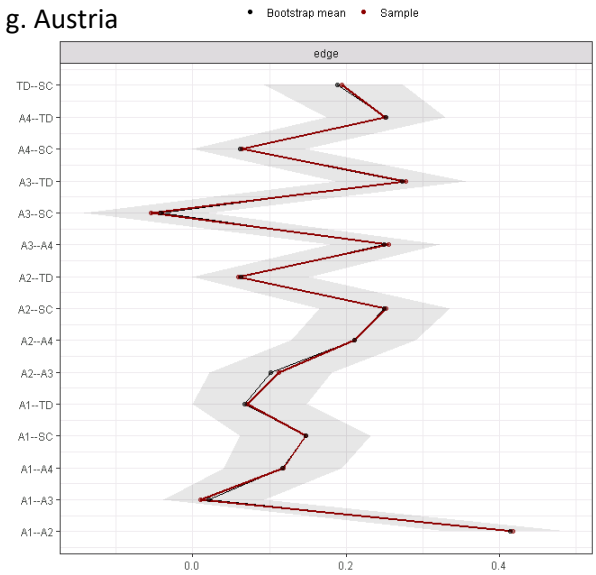
f. US



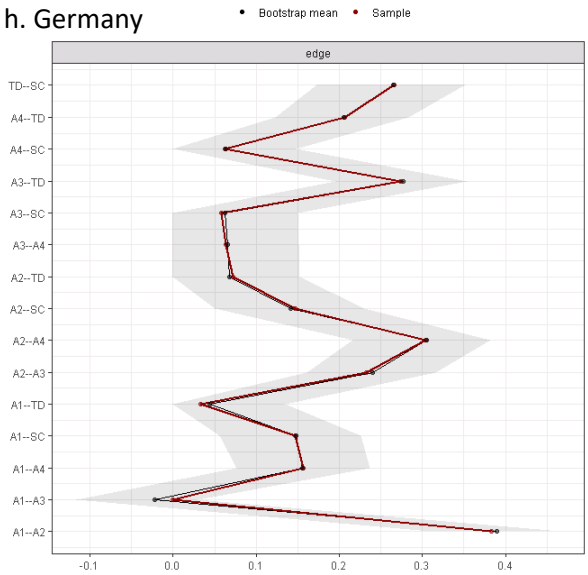
Supplement 1 (continued)

Edge Recovery across countries - Bootstrapping

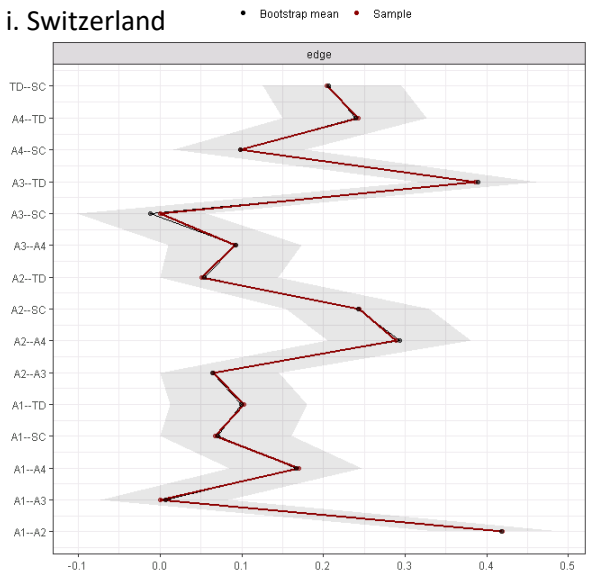
g. Austria



h. Germany



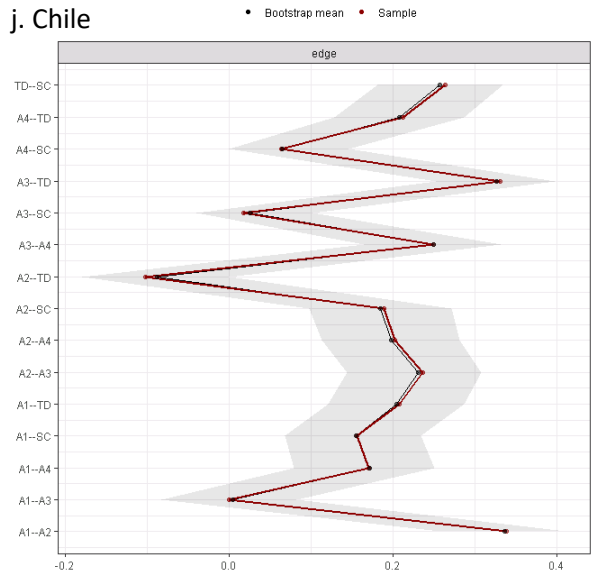
i. Switzerland



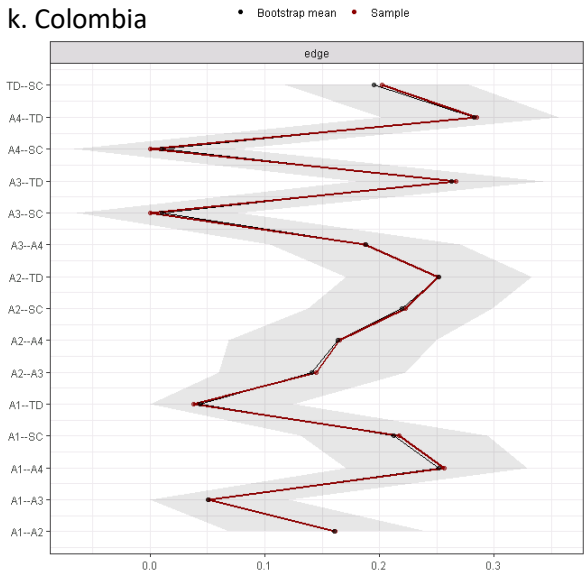
Supplement 1 (continued)

Edge Recovery across countries - Bootstrapping

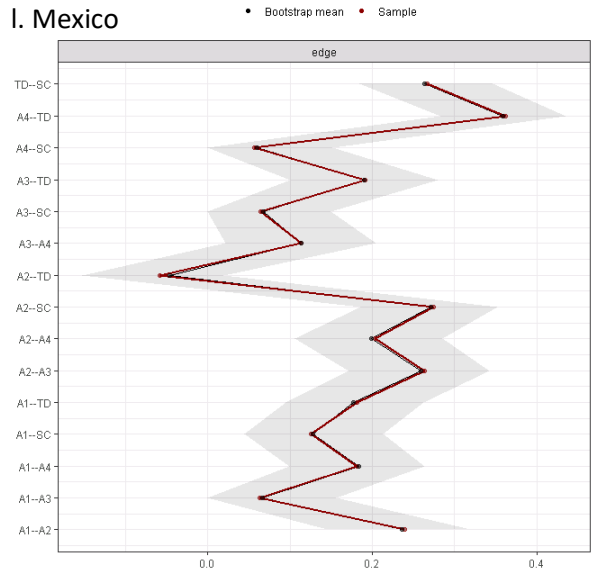
j. Chile



k. Colombia



l. Mexico



Note. A1= feedback performance in class, A2= feedback individual strength and weaknesses, A3= informing about expectations in test, A4= feedback how to improve, TD= rounded mean score for teacher-directed practices, SC=rounded mean score for student-centred practices.

Supplement 2

The CS (cor=0.7) Coefficient for the Accuracy and Stability of Strength Centrality across Countries - Bootstrapping

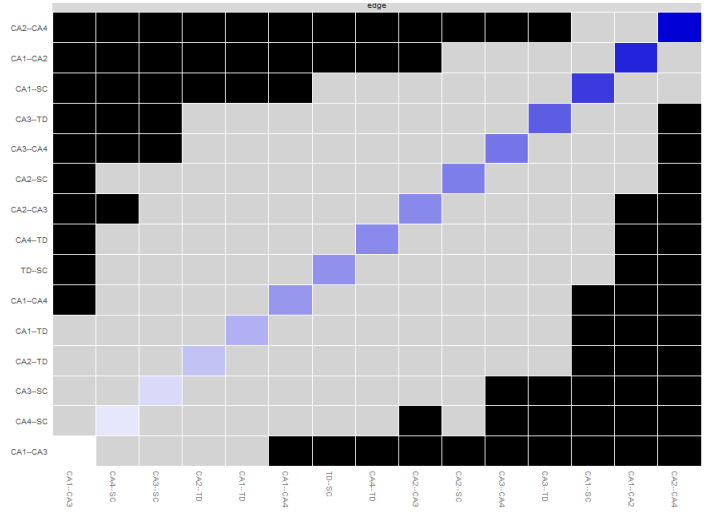
Country	Strength
Macao	0.67
Taipei	0.52
Shanghai	0.67
Australia	0.75
UK	0.67
US	0.44
Austria	0.52
Germany	0.60
Switzerland	0.75
Chile	0.36
Colombia	0.67
Mexico	0.52

Note. Values <0.5 indicate network stability and accuracy.

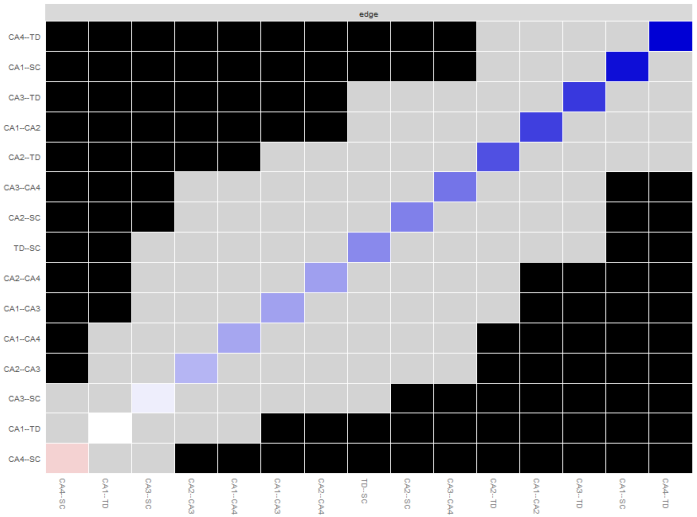
Supplement 3

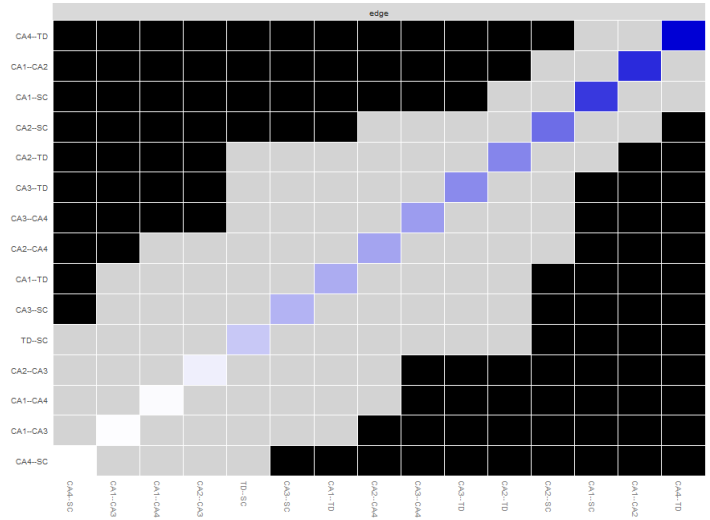
Bootstrapped Differences between all Edge Weights across Countries

a. Macao



b. Taipei



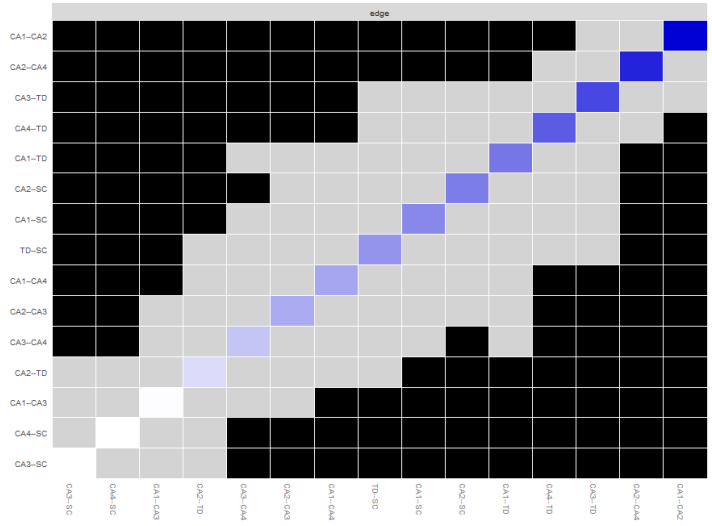


c. Shanghai

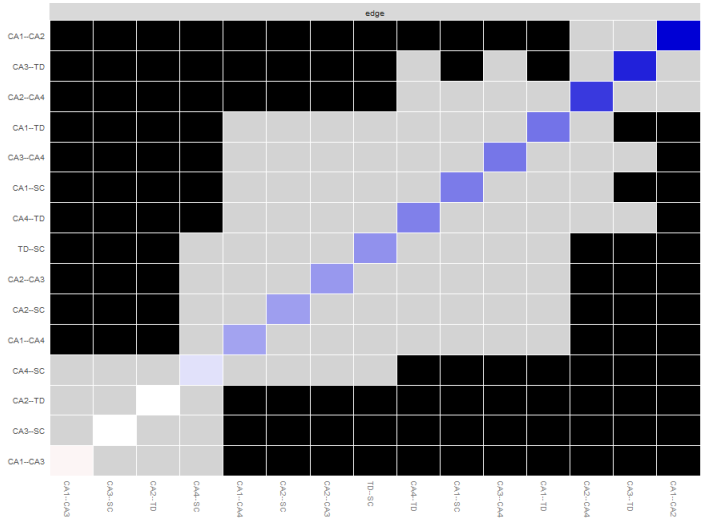
Supplement 3 (continued)

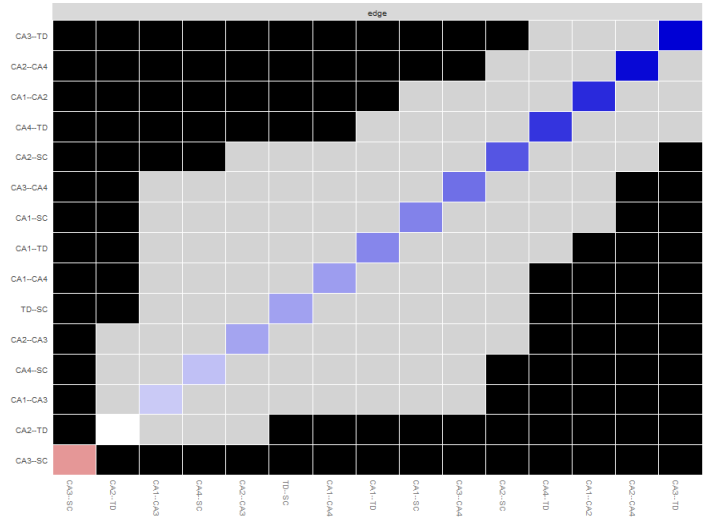
Bootstrapped Differences between all Edge Weights across Countries

d. Australia



e. UK



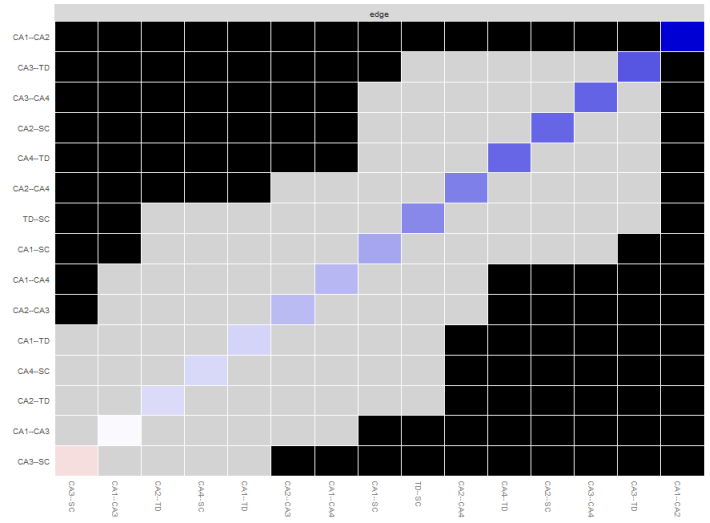


f. US

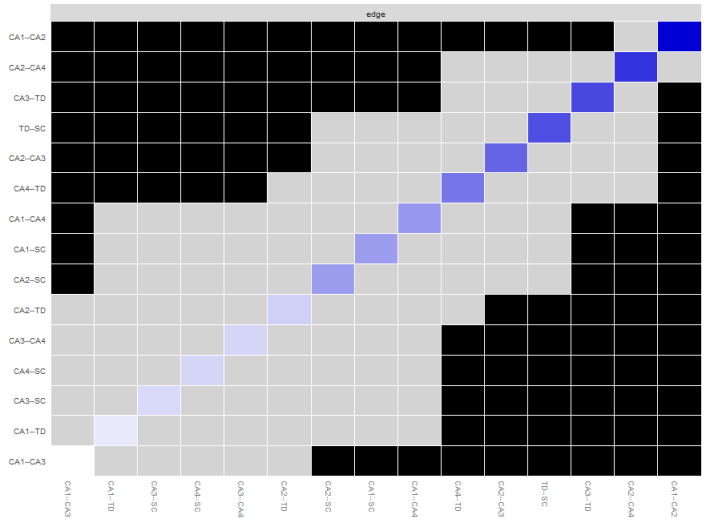
Supplement 3 (continued)

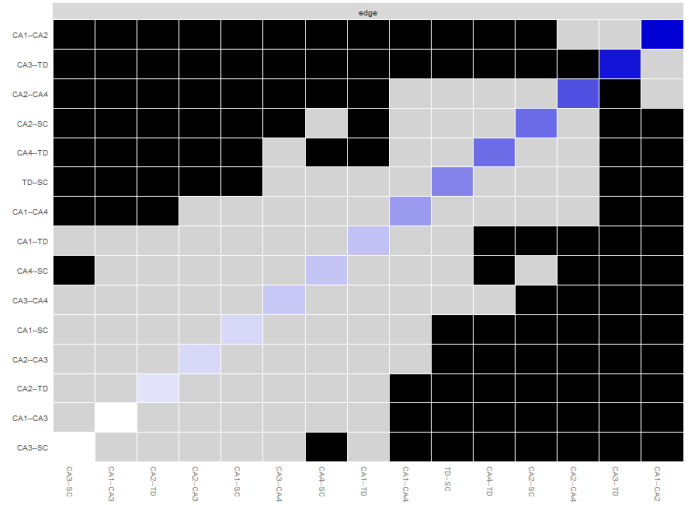
Bootstrapped Differences between all Edge Weights across Countries

g. Austria



h. Germany



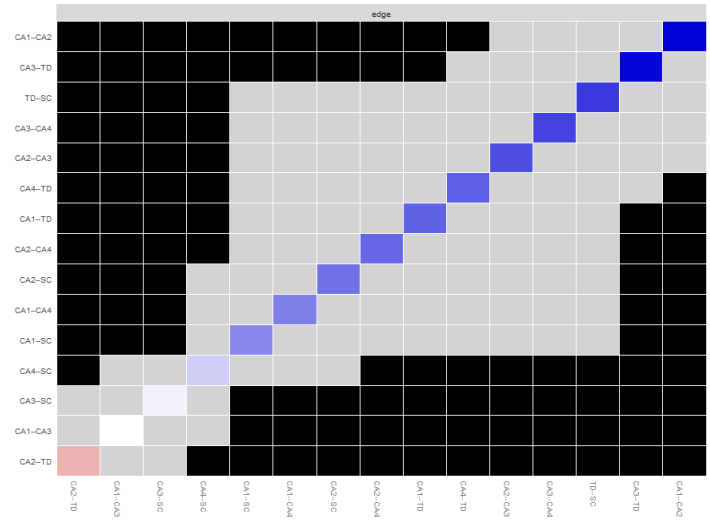


i. Switzerland

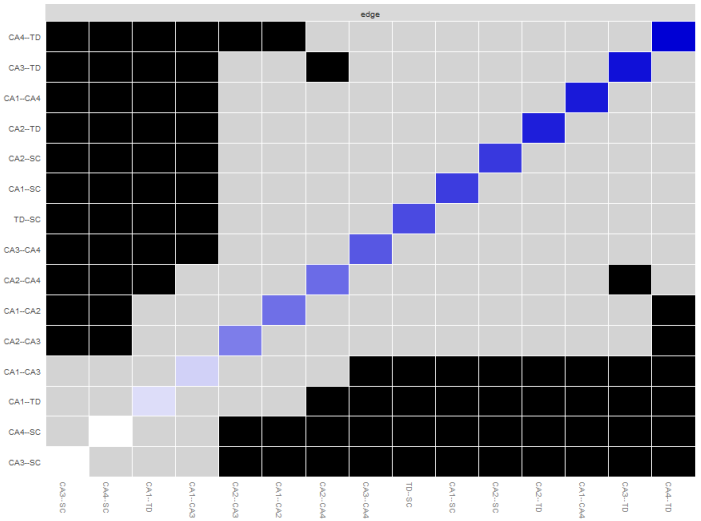
Supplement 3 (continued)

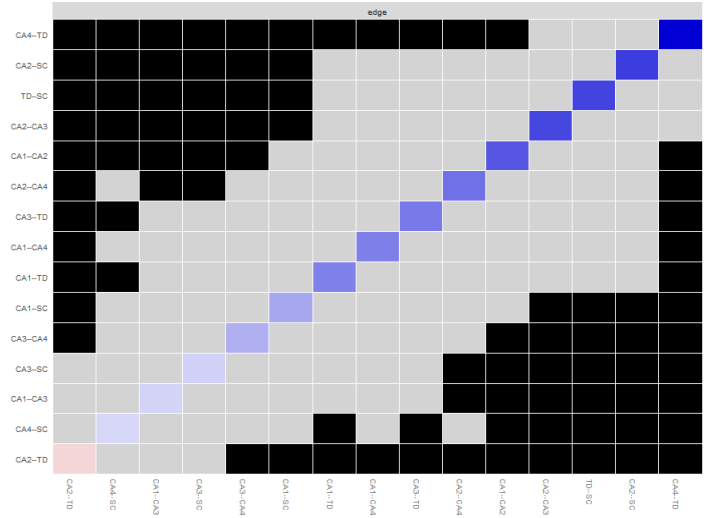
Bootstrapped Differences between all Edge Weights across Countries

j. Chile



k. Colombia





l. Mexico

Note. CA1= feedback performance in class, CA2= feedback individual strength and weaknesses, CA3= informing about expectations in test, CA4= feedback how to improve, TD= rounded mean score for teacher-directed practices, SC=rounded mean score for student-centred practices.

Supplement 4

Individual Node Strength-centrality across Countries

	A1	A2	A3	A4	TD	SC
AUS	0.92	1.10	0.50	0.81	0.96	0.54
AUT	0.76	1.05	0.71	0.90	0.85	0.71
CHE	0.76	1.07	0.54	0.89	0.99	0.61
CHL	0.87	1.07	0.83	0.90	1.12	0.69
COL	0.72	0.95	0.65	0.89	1.04	0.64
GER	0.72	1.14	0.63	0.79	0.85	0.68
UK	0.96	1.00	0.72	0.90	0.92	0.56
MAC	0.82	1.07	0.62	0.87	0.72	0.68
MEX	0.79	1.04	0.70	0.92	1.06	0.79
QCN	0.83	0.97	0.51	0.74	1.04	0.78
TAP	0.87	0.93	0.74	0.88	1.05	0.77
US	0.87	1.02	0.92	1.07	0.95	0.79
Average across all countries	0.82	1.03	0.67	0.88	0.96	0.69

Note. AUS= Australia, AUT=Austria, CHE=Switzerland, CHL=Chile, COL=Colombia, GER=Germany, UK= Great Britain, MAC= Macao, MEX=Mexico, QCN= Shanghai, TAP= Taipei, US= United States. A1= feedback performance in class, A2= feedback individual strength and weaknesses, A3= informing about expectations in test, A4= feedback how to improve, TD= rounded mean score for teacher-directed practices, SC=rounded mean score for student-centred practices.