Comparability of perceived feedback across culturally different nations

Jitske de Vries
University of Twente

Graduation committee

Prof. dr. ir. Bernard Veldkamp
(University of Twente)

Prof. dr. ir. Theo Eggen
(University of Twente)

Dr. Remco Feskens
(Cito)

Dr. Jos Keuning
(Cito)

July 2018

Abstract

Feedback in the classroom is an important factor in increasing educational outcomes of students. One of the studies that takes feedback into account in its surveys is the Programme for International Student Assessment (PISA). PISA is administered in as many as 73 countries all over the world. Research, however, indicates that what feedback actually means might differ across culturally different nations. An important characteristic of culture is a nation's tendency to be either more individualistic or collectivistic oriented. This difference in meaning might result in a decrease in usability of the feedback scale, as comparisons across countries are less meaningful. This research aimed to investigate the feasibility of the feedback scale as currently posed by PISA through (1) investigating whether measurement invariance holds across culturally different nations, and (2) assessing the impact of model fit of the feedback scale with regard to educational outcomes. The results of the first research question showed that there are indications that measurement invariance does not completely hold for the feedback scale across culturally different nations. However, the analysis of the impact in the second research question showed that the current feedback scale can be used regardless of the inability to hold measurement invariance. For future measurement of feedback, it is advisable to replace some questions of the scale in order to be able to measure broader aspects of feedback. Further investigations into the measurement of feedback across cultural countries are needed to assess the missing aspects of feedback in the current scale.

Keywords: feedback, measurement invariance, culture

Table of Contents

1. Introduction

Feedback in the classroom is a crucial factor in increasing educational outcomes of students (Shepard, 2006; Black & Wiliam, 2004; McMillan, 2007). The reason for this is that feedback has the potential to bridge the gap between where students are and where they are going and to indicate what is needed to bridge this gap (Hattie & Timperley, 2007). Feedback is found to be a complex subject as there are various definitions and aspects of feedback. Also, the specific content and situation of feedback determine its power. For example, multiple studies have shown that more elaborate feedback can be preferred above short, verification feedback (Hattie & Timperley, 2007; Shute, 2008). The power of feedback is mostly dependent on school, teacher and learner characteristics. One of the learner characteristics is culture (Hattie, 2009). For example, De Luque and Sommer (2000) showed that students in collectivistic countries prefer implicit feedback as opposed to individualistic countries where explicit feedback is the standard.

The variation of implementation and impact of feedback across different countries are the causes of the fact that feedback is often measured as background construct in cross-national studies (OECD, 2017). Such cross-national studies, which are called International Large-Scale Assessments (ILSA), are focused on mapping student performance on various subjects. One of the ILSAs that takes feedback into account in its surveys is the Programme for International Student Assessment (PISA). Every three years, PISA assesses the level of 15-year-old students in various subjects, most importantly science, reading and mathematics (OECD, 2017). Moreover, PISA distributes background questionnaires to the students, parents and teachers to gain information on the (educational) background of students. In the student background questionnaire, PISA added the measurement of the construct 'Perceived feedback' in order to assess whether students perceive feedback at all (Bayer, Klieme, & Jude, 2016). The difference between perceived feedback and given feedback lies within the characteristics of the student (Hattie & Timperley, 2007). For example, feedback given at an inappropriate moment or at an inappropriate level for the particular student might be discarded by the student.

PISA uses various questions to measure the perceived feedback of students. However, it is uncertain whether feedback has the same meaning across culturally different nations. One reason for this uncertainty is the counterintuitive negative relationship between the latent variable 'Perceived feedback' and other external variables, such as educational outcomes (OECD, 2016). It is possible that feedback has a different meaning and interpretation in

different countries and that this might lead to a case where measurements are not invariant with respect to country or culture. According to Mellenbergh (as cited in Van de Schoot, Schmidt, De Beuckelaer, Lek, & Zondervan-Zwijnenburg, 2015), "…measurement invariance requires that the association between the items (or test) scores and the latent factors (or latent traits) of individuals should not depend on group membership or measurement occasion (i.e., time)." When this assumption does not hold it is less meaningful to make comparisons across groups such as countries or cultures (Schmitt & Kuljanin, 2008). Furthermore, such comparisons easily lead to wrongful conclusions (Steenkamp & Baumgartner, 1998). An urgent need thus exists for research on the comparability of the relationships (i.e., measurement invariance) between 'Perceived feedback'-scores across several PISA 2015 nations. This study aims to investigate the feasibility of the 'Perceived feedback'-scale by assessing (1) whether the assumption of measurement invariance holds across culturally different nations, and (2) the impact of model fit of the feedback scale in relation to educational outcomes.

2. Conceptual framework

2.1. Definition of feedback

Feedback is an often researched topic in the educational field and many have tried to define it (Shute, 2008; Hattie & Timperley, 2007; Black & Wiliam, 1998; Kluger & DeNisi, 1996) According to Ramparsad (1987), feedback can be defined as "… information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way." So, feedback is not just information, but information to serve particular instructional purposes and it has to have an impact on future performance. This is why feedback is often domain-specific. Moreover, Hattie and Timperley (2007) describe feedback as "…information provided by an agent (e.g., teacher, peer, book, parent, and experience) regarding aspects of one's performance or understanding." According to their study, feedback can come in multiple levels and types, as is illustrated in Figure 1 (reprinted from Hatty and Timperley, 2007).

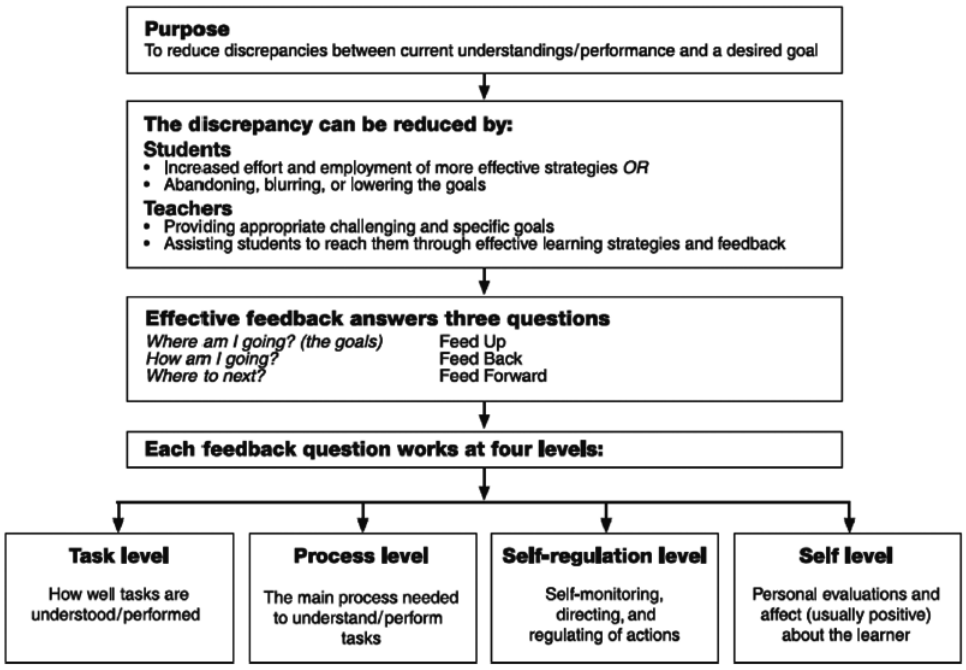

Figure 1. Hattie and Timperley's model of feedback (2007).

As can be seen in Figure 1, Hattie and Timperley (2007) describe three types of feedback: Feed Up, Feed Back, and Feed Forward. In Feed Up, the student gets feedback that is related to a clearly defined goal. This goal should be appropriately challenging for the student, because it leads to clarity on which actions should be undertaken on the part of the student and the teacher. In Feed Back, the students get feedback which shows them how they are doing in relation to the set goal. In Feed Forward, the students get feedback on how they can

deepen and broaden their knowledge. Hattie and Timperley (2007) further mention four levels at which feedback can be given, namely at the task level (FT), process level (FP), self-regulation level (FR) and self-level (FS). First, feedback can be related to how the learner does the task (FT). Second, feedback can be directed at how the learner processes the information (FP). Third, feedback can give the learner information about how to reflect on a skill or gain confidence in a certain task (FR). Fourth, feedback can be given about the learner personally (FS). The effectiveness of these four levels differ greatly. Feedback at the task level is best (FT), when the purpose is to increase effective strategizing. In contrast, feedback on the process (FP) and self-regulation (FR) level are best when deep processing and mastery is pursued. Almost all research concurs that feedback on the level of the self (FS) is least beneficial (Butler, 1987; Kluger & DeNisi, 1996).

Feedback can be translated to the classroom in the form of formative assessment. Formative assessment can be defined as "… information communicated to the learner that is intended to modify his or her thinking or behaviour for the purpose of improving learning." (Shute, 2008). In her study, Shute (2008) elaborates even more on the aspects of feedback and mentions the importance of the direction of feedback (i.e., positive or negative), the elaborateness of the feedback, and timing (i.e., immediate or delayed). The student characteristics are important in deciding the type and level of feedback (Shute, 2008; Hattie & Timperley, 2007). Among these characteristics is the culture of the student (Hattie & Timperley, 2007).

2.2. Perceived feedback across cultural dimensions

The difference between cultures of countries is often characterized by the individualism-collectivism dimension as posed by Hofstede (1986). The validity of this dimension has been demonstrated by multiple studies (Oyserman, Coon, & Kemmelmeier, 2002; Hofstede et al., 2010). Hofstede, Hofstede and Minkov (2010) define individualism as "…a preference for a loosely-knit social framework in which individuals are expected to take care of only themselves and their immediate families." Hofstede et al. (2010) define collectivism as "…a preference for a tightly-knit framework in society in which individuals can expect their relatives or members of a particular ingroup to look after them in exchange for unquestioning loyalty." Several of the most important differences between individualistic and collectivistic countries as they are established by Hofstede (2011) can be found in Table 1.

Table 1. *Several differences between individualism and collectivism (Adapted from Hofstede, 2011).*

Collectivism	**Individualism**
"We" – consciousness	"I" – consciousness
Harmony should always be maintained	Speaking one's mind is healthy
Others classified as in-group or out-group	Others classified as individuals
Languages in which the word "I" is avoided	Languages in which the word "I" is indispensable
Purpose of education is learning how to do	Purpose of education is learning how to learn
Relationship prevails over task	Task prevails over relationship

Instructional practices differ in several ways across countries with either an individualism or collectivism tendency (Hofstede et al., 2010; Kaur & Noman, 2015). Teachers in an individualistic country use a more student-centred approach: students are motivated by their teacher to become independent thinkers and to take initiative (Prosser & Trigwell, 1998; Staub & Stern, 2002; Faitar, 2006; Cothran et al., 2005). This means that students are expected to speak up when they have questions or when questions are asked by the teacher (Hofstede et al., 2010). Teachers also have a certain amount of autonomy in their lessons and can therefore also spend more attention to individual differences (Lau & Lam, 2017). Teachers in a collectivistic country take a more teacher-centred approach: the teacher transfers the knowledge to the passive students (Prosser & Trigwell, 1998; Staub & Stern, 2002). The teacher guides the student each step of the process (Cothran et al., 2005). It is also expected that both teacher and student will do their best to retain their good status (Hofstede et al, 2010). In order to do this, a strict hierarchy is maintained in which the student will only respond when he is personally called upon by the teacher. Furthermore, teachers emphasize the importance of the students' effort and therefore adaptive instruction (and feedback) seem to be less linked to performance in collectivistic countries (Lau & Lam, 2017). The lessons of the collectivistic teacher are also restricted by a greater importance on examinations (i.e., teaching to the test) (Wang, 2013; Bai, 2010).

However, differences in pedagogical approaches are not only found between collectivistic and individualistic countries, but also within these countries. This is evident in the study of Lau and Lam (2017), who researched the pedagogical approaches of the top-10 performing PISA countries. Teacher-directed teaching is common in collectivistic countries, but it is also used in individualistic countries Canada and Finland, whereas it less common in collectivistic countries Japan and Korea. Notably, China, as one of the most collectivistic countries, uses a variety of pedagogical approaches to reach the student and not only the teacher-centred approach. Adaptive teaching is mostly common in collectivistic countries, but

it is also used in Canada. According to Lau and Lam (2017), one must therefore keep in mind that culture is not necessarily homogeneous and stable, but rather flexible and dynamic.

The practice of feedback in the classroom may also differ with regard to individualism-collectivism dimension. According to De Luque and Sommer (2000), learners in collectivistic cultures prefer to get group-focused feedback, which is implicit and preferably not related to the individual. The teacher in a collectivistic country will deal with the students as a group (Suhoyo et al., 2014). However, Lau and Lam (2017) noticed that some students in the more collectivistic countries China, Singapore and Chinese Taipei actually prefer more individual feedback. Learners in individualistic countries, however, are mostly interested in feedback that is direct and related to the individual effort. Furthermore, learners in individualistic countries also are more likely to seek feedback themselves and, thus, be more independent of the teacher.

The preference for the source of feedback also differs between individualistic and collectivistic countries. According to Hwang and Fransesco (2010), learners in collectivistic countries seem to prefer to get feedback from peers, such as classmates or group members, instead of teachers. The reason for this is that they see their classmates as equals, while their professor is seen as superior. The power distance between the students and the professor is often an important factor of feedback for learners in collectivistic cultures. A greater power distance also leads to more one-way communication and it is not expected of students to seek feedback (Suhoyo et al., 2014). Especially in Japan and Korea, two more collectivistic countries, students are not given opportunities to ask for feedback (Lau & Lam, 2017). Furthermore, learners in individualistic countries do not seem to have this struggle with hierarchical differences (i.e., the power distance). Instead, these learners prefer to get feedback from professors as these are seen as more knowledgeable than themselves and thus more likely to contribute to their own knowledge. In the case of a smaller power distance, two-way communication is the norm and teachers expect that students seek feedback (Suhoyo et al., 2014).

The purpose of seeking and receiving feedback is also different across individualistic and collectivistic countries. According to Kung (2008), learners in both individualistic and collectivistic countries are interested in feedback for its function to reduce uncertainty. However, learners in collectivistic countries tended to appreciate developmental and critical feedback, because the collective could also benefit from it. In the latter, learners mostly

sought feedback that helped them to look better and to protect their egos (Brutus & Greguras, 2008).

2.3. Measurement invariance and comparability

Perceived feedback cannot be measured directly. Instead it is a latent variable which is measured by some observable variables. ILSAs often administer questionnaires with these observable variables in order to be able to measure latent variables. The scores on the latent variable are thus the result of underlying relations between the (observed) variables. In the case of perceived feedback, the score is based on a measurement model consisting of 5 questions. The resulting scores are often used to indicate causal relationships between educational practices and educational outcomes. However, such analysis can only be done if the latent variable has the same meaning for each group. Latent scores can be distorted if the underlying scale does not have the same meaning across different groups. For example, in the study by Täht and Must (2013), it was assessed whether the learning attitudes of students across different nations had the same meaning across culturally different nations. In this example, the scale of learning attitudes did not measure the same underlying (latent) construct across countries. In such cases the assumption of measurement invariance does not hold and wrongful conclusions can easily be made (Steenkamp & Baumgartner, 1998). When the assumption of measurement invariance across culturally different nations does not hold, this can be caused by the influence of cultural values and educational practices of the student's country (Täht & Must, 2013). For example, questions can be interpreted differently or a latent construct can have a different meaning. As mentioned, it is very possible that the aspects of feedback that are considered important depend on the cultural characteristics of the country.

2.4. Effect of feedback on educational outcomes

As mentioned, different levels and types of feedback can be found in the classroom. Nyquist (2003) made a typology of these different kinds (Table 2). He noticed that whenever the teacher gave the student more elaborate feedback, this lead to higher effect sizes on educational outcomes.

Table 2. *Effective formative assessments (Reprinted from Nyquist, 2003).*

Type	Description
Weaker feedback only	Students are given only the knowledge of their own score or grade, often described as "knowledge of results."
Feedback only	Students are given their own score or grade, together with either clear goals to work towards, or feedback on the correct answers to the questions they attempt, often described as ''knowledge of correct results.''
Weak formative assessment	Students are given information about the correct results, together with some explanation.
Moderate formative assessment	Students are given information about the correct results, some explanation, and some specific suggestions for improvement.
Strong formative assessment	Students are given information about the correct results, some explanation, and specific activities to undertake in order to improve.

Regardless of these effect sizes, it is not the case that teachers mostly give strong formative feedback. Black and Wiliam (2004) mention that in 38% of the classrooms, teachers tend to give feedback that is focused on the self (i.e., praise). Also, feedback on the task is in 90% of the cases combined with feedback on the self, which dilutes the power of feedback on the task (Hattie & Timperley, 2007). The reason for this is that feedback at level of the self is often not informative and does not lead to more engagement of the student. There is, however, evidence that praise of the individual can motivate the learner to learn in the face of success. Nevertheless, when such a learner is subsequently faced with bad news, the learner might take this failure too personally which is again detrimental for his learning process (Henderlong Corpus & Lepper, 2007). It is therefore very possible that studies that did not found a positive effect of feedback, were because of the negative effect of feedback on the self.

Hattie and Timperley (2007) mention that a learner has four possible ways to deal with feedback: increased effort/use of better strategies, abandoning of the goals, blurring of the goals, and lowering of the goals. Hattie (2009) mentions the importance of understanding how feedback is received in the classroom in order for feedback to be beneficial. Shute (2008) also elaborates on this subject by establishing a long list of types of situations and corresponding level and types of situations. For example, teachers should not make comparisons between students as this draws attention to the "self" and teachers should not

give elaborate and immediate feedback to a high performing student instead of a lower performing student (Shute, 2008; Hattie & Timperley, 2007). It is also very possible that the type and level of feedback are dependent on the culture of the student, because, as mentioned previously, the practice of feedback varies greatly across culturally different nations.

2.5. Perceived feedback as measurement in International Large-Scale Assessments

In order to measure feedback as a background variable of students cross-nationally, PISA developed a scale consisting of five questions that are meant to measure 'Perceived feedback'. The difference between perceived feedback and feedback often lies with the student and his characteristics (Hattie & Timperley, 2007). The teacher can give a student much feedback, but if a student does not perceive it as such, it has no power. Hattie and Timperley (2007) mention that not recognizing feedback is one of the pitfalls of the effectiveness of feedback. Herein also lies the danger of self-reporting. For example, a student may lack the introspective ability to accurately detect feedback or students lack understanding or have a different understanding of what feedback actually is. Also, PISA struggles with missing values on the background questionnaire and also on the 'Perceived feedback'-scale. For example, of the values on the 'Perceived feedback'-scale, approximately 15% is missing.

The reliability of the questionnaire is assessed by an expert group on several quality criteria: psychometric properties, range of difficulty, durability, and interest (OECD, 2016). Several quality assurance mechanisms are in place as to retain the validity and reliability as translation, sampling and data collection take place (OECD, 2016). Interestingly, comparative studies that have tested the effect of feedback on the performance of students have found counterintuitive results, namely they found a negative correlation between 'Perceived feedback' and student performance (OECD, 2016). One way to explain this is that as students become more skilled, the amount of feedback they receive from the teacher reduces. Another way to explain this result is that the established 'Perceived feedback'-scale might be incomplete according to the mentioned feedback models or that 'Perceived feedback' is something different across different cultures.

2.6. Research question and model

One of the most important goals of ILSAs, such as PISA, is to map student achievement and background factors of education across international countries. One of these contextual factors in PISA is 'Perceived feedback' and this factor is measured, because of its great

impact on student performance. As mentioned, what feedback is and when it is perceived as such, might be culturally defined. To be able to assess this relationship meaningfully, the measurement of the scale must be invariant across groups. Due to the reason that the meaning of feedback might differ per culturally different country, the 'Perceived feedback'-scale might not be invariant and this has implications for the interpretation of the scale. This can have an impact on current educational practices, for example, countries that want to increase on their feedback practices in the classroom. Moreover, it can influence the established research about effective feedback across cultures.

This research will investigate two aspects with regard to the feasibility of the current 'Perceived feedback'-scale as measured within PISA 2015. The first research question will assess whether the assumption of measurement invariance across culturally different nations holds for the 'Perceived feedback'-scale. The second research question will assess the impact of culturally different nations on the relationship between 'Perceived feedback' and educational outcomes. This resulted in the following research question:

'To what extent is perceived feedback comparable across culturally different nations?'

With two sub-questions:

1. *'To what extent are the scores on 'Perceived feedback' subject to measurement invariance (i.e., comparable) across various PISA 2015 nations?'*
2. *'What is the impact of culturally different nations on the relationship between 'Perceived feedback' and educational outcomes?'*

3. Research design and methods

3.1. Research design

This research has a quantitative nature: the answers of the research questions were sought with the use of the database of PISA 2015 and Structual Equation Modeling (SEM). The first research question entailed analyses of the measurement model of the 'Perceived feedback'-scale. This model was analysed with regard to measurement invariance, which analysis is based on the checklist of Van de Schoot, Lugtig and Hox (2012) and which makes use of Confirmatory Factor Analysis (CFA). A first step was to make a categorisation of PISA 2015 nations as either 'individualistic' or 'collectivistic'. After that it was evaluated whether the assumption of measurement invariance holds across these two groups of countries. This was done with several measurement models, each with a different set of constraints on factor loadings, intercepts, residuals. With the results of these models it could be ascertained whether there was a case of measurement invariance. The impact of a possible case of measurement invariance was established by answering the second research question. To answer the second research question, SEM, which combines a measurement and structural part, was applied as well (Kline, 2015). First, a model was estimated in which the latent variable 'Perceived feedback' was linked to the scores of one of the proficiency tests. Second, a model was estimated in which the latent variable was estimated separately for the groups of countries (i.e., 'individualistic' vs. 'collectivistic'). The difference between these models identified the cultural impact on the relationship between 'Perceived feedback' and educational outcomes.

3.2. Respondents

Both research questions were answered with the use of the data of PISA 2015. The general sample of the PISA 2015 survey consists of approximately 540 000 students, representing about 29 million 15-year-olds in the schools of the 72 participating countries. These participants were sampled by PISA by using a two-stage stratified sampling method. In the first stage, a random sample of 150 schools are drawn from the total list of schools. In the second stage, from each school, 35 15-year-old students are randomly selected to participate. For more details on the methods for sampling, see the PISA 2015 technical report (OECD, 2017).

To assess difference in feedback between countries that differ on the individualism-collectivism dimension, the PISA countries were categorised. This categorisation was based

on the values that PISA countries have on the Individualism index (IDV) in the study of Hofstede et al. (2010). For the placement of all the PISA countries on the IDV, see Appendix A. The PISA countries that fall into the lowest quartile (≤27) are considered as collectivistic and the PISA countries that fall into the highest quartile (≥68) are considered to be individualistic. These cut-off scores were chosen as they approximately represented the 25% lowest scoring countries (i.e., the most collectivistic countries) and the 25% highest scoring countries (i.e., the most individualistic countries). Ten PISA countries were excluded from the categorisation as they were not included in the study by Hofstede et al. (2010) or did not have data for the 'Perceived feedback'-scale. This eventually led to the selection of sixteen collectivistic countries and sixteen individualistic countries (Table 3). All of the participants in these countries are used in the analyses.

Table 3. *Categorisation of collectivistic and individualistic countries.*

Collectivistic countries	**N**	**IDV score**	**Individualistic countries**	**N**	**IDV score**
Portugal	7325	27	U.S.A.	5712	91
Puerto Rico	1398	27	Australia	14530	90
Hong Kong	5359	25	U.K.	14157	89
U.A.E.	14167	25	Canada	20058	80
Chile	7053	23	Netherlands	5385	80
Singapore	6115	20	Hungary	5658	80
Vietnam	5826	20	New Zealand	4520	79
Thailand	8249	20	Italy	11583	76
B-S-J-G (China)	9841	20	Belgium	9651	75
Korea	5581	18	Denmark	7161	74
Taiwan	7708	17	France	6108	71
Peru	6972	16	Sweden	5458	71
Trinidad and Tobago	4692	16	Ireland	5741	70
Costa Rica	6866	15	Latvia	4869	70
Colombia	11795	14	Norway	5456	69
Indonesia	6513	13	Switzerland	5860	68
Total	115460		**Total**	131907	

3.3. Instrumentation

The scores on the 'Perceived feedback'-scale of the background questionnaire and the scores of the various proficiency tests of PISA 2015 were used order to assess measurement invariance and the possible impact of the failure to hold measurement invariance. To generate the scores on the measurement models of the 'Perceived feedback'-scale and the proficiency

tests, item response theory (IRT) methodology was used (Kaplan & Kuger, 2016). In this method, observable item responses are linked to underlying latent trait.

Perceived feedback.

The 'Perceived feedback'-scale as established by PISA 2015 consists of five Likert scale items, which were all developed to measure whether (formative) feedback in the classroom perceived at all (Bayer, Klieme, & Jude, 2016). These items were:

- "The teacher tells me how I am performing in this course." (ST104Q01NA)
- "The teacher gives me feedback on my strengths in this science subject." (ST104Q02NA)
- "The teacher tells me in which areas I can still improve." (ST104Q03NA)
- "The teacher tells me how I can improve my performance." (ST104Q04NA)
- "The teacher advises me on how to reach my learning goals." (ST104Q05NA)

These items were rated on a 4-point Likert scale: 'Every class', 'Most classes', 'Some classes', and 'Never or hardly ever'. To scale these Likert scale items, the partial-credit model and the classical test theory were used. This model resulted in weighted likelihood estimates that were used as 'Perceived feedback'-scores for the students. In order to facilitate interpretation, these scale indices were transformed to an international mean of zero and standard deviation of one (Kaplan & Kuger, 2016; OECD, 2017).

Math, reading and science proficiency.

To assess the impact of the failure to hold measurement invariance, the scores on the scientific, reading and mathematical proficiency tests as administered by PISA 2015 were used. The scientific items aimed to assess content, procedural and epistemic knowledge and several scientific competencies. The reading items aimed to assess whether learners could access and retrieve information, integrate and interpret on information, and reflect and evaluate on information. The mathematical items aimed to assess three mathematical processes of different mathematical contents and contexts: formulating situations mathematically, employing mathematical concepts, and interpreting, applying and evaluating mathematical outcomes (OECD, 2017). For more information on the items, see the PISA 2015 manual (OECD, 2017). To take the measurement uncertainty into account, PISA makes use of the plausible values methodology. The second research question also used these plausible values in order to estimate the scores of the students on the math, reading, and

science proficiency tests. In PISA 2015, 10 plausible values were drawn from the posterior distribution for each subject. The plausible values served as ability estimates and provide information about the uncertainty of these estimates. In order to facilitate interpretation, the plausible values were transformed to a scale with an international mean of 500 and a standard deviation of 100 (OECD, 2017).

3.4. Procedure

The answering of both research questions required the data of PISA 2015 and various statistical analyses. The data from the PISA 2015 tests are downloaded from the OECD website (http://www.oecd.org/pisa/data/2015database/). The statistical analyses of the data was done with the use of *R* and *Rstudio* (R Development Core Team, 2015; Rstudio Team, 2015), which are suitable all SEM analyses. The packages that were used for the SEM analyses are *lavaan* (Rosseel et al., 2017) and *lmtest* (Hothorn, 2018). For the graphical illustrations of the distribution of mean scores, the package *ggplot2* was used (Wickham, 2016). Only the illustrations of the measurement and structural models was done with *Microsoft PowerPoint*. For the second research question, the proficiency scores, which are based on plausible values were also needed. These scores were calculated with the use of the *intsvy* package (Caro & Biecek, 2017).

3.5. Data analysis

Before analysing the data, the data was screened to see whether there are outliers present in either of the groups. The reason for this is that outliers induce bias and this influences the factor models (Van de Schoot, Lugtig, & Hox, 2012).

The first research question entailed an investigation towards the assumption of measurement invariance. First, the descriptive statistics were calculated (i.e. mean and standard deviation per country on the items of the scale and total score). These descriptive statistics are graphically illustrated. Reliability was assessed with the use of Cronbach's Alpha, which was calculated for the collectivistic countries, individualistic countries, and for each country separately.

The analysis of measurement invariance makes use of CFA, which is a way to model the relationship of observational variables with underlying latent variables (Gallagher & Brown, 2013). Before starting to analyse measurement invariance, the model that was used by PISA 2015 was analysed. This model is also the strictest model, because it is assumed that measurement invariance is established. Then, the analysis of strict measurement invariance

(Meredith, 1993) of 'Perceived feedback'-scale was carried out based on the checklist by Van de Schoot, Lugtig and Hox (2012) and consisted of four steps. In each step, the amount of constraints on the model was changed and/or reduced. Thus, we started with the strictest model and we released a constraint in each step.

1. Test whether the explained variance for every item is identical across groups. The factor loadings are correlations between the items and the construct and indicate whether differences among responses are caused by the level of the construct (i.e., the items have the same meaning for each respondent). The intercept is the starting point of the scale of the construct and participants that have the same value on the construct should also have the same values on the items of the construct. The residuals reflect the overall error in the prediction that is based on the indicator variables and these should be the same across groups and across times (Byrne, 2012). This entails a model where the intercepts, factor loadings, and residuals are constrained to be equal across groups.
2. Test whether the meaning of the construct (factor loadings), and the meaning of the levels of the underlying items (intercepts) are equal in both groups (scalar invariance). This entails a model where both intercepts and factor loadings are constrained to be equal across groups.
3. Test whether tests the meaning of the levels of the underlying items (intercepts) are equal in both groups (intercept invariance). This entails a model where only the intercepts are constrained to be equal across groups, but where factor loadings are not constrained.
4. Test whether respondents across groups attribute the same meaning to the latent construct under study (metric invariance). This entails a model where the factor loadings are constrained to be equal across the groups, but where the intercepts are not constrained. For all the equality constrains of the model parameters, see Table 4.

Table 4. Equality constraint of model parameters (based on Täht and Must, 2013).

Model	Factor loadings	Intercepts	Residuals
1 (Strict MI)	Invariant	Invariant	Invariant
2 (Scalar invariance)	Invariant	Invariant	Free
3 (Intercept invariance)	Free	Invariant	Free
4 (Metric invariance)	Invariant	Free	Free

The models were compared with each other using Likelihood Ratio tests, which compare the significance of the difference of the Chi-square value between the models. The significance determined whether the model fit is worse than the previous one and, therefore, whether the assumption of measurement invariance can hold. After conducting the CFA's and establishing whether the 'Perceived feedback'-scale is measurement invariant, the impact of the culturally different nations was further analysed by two models in which 'Perceived feedback' was estimated separately for both collectivistic and individualistic countries. Within each group of countries, the model was further analysed by looking at the fit of the model for each separate country. The country for which the goodness-of-fit indices showed the most appropriate levels, is illustrated. Finally, the modification indices of the current 'Perceived feedback'-scale were evaluated. The impact of the modifications to the model of the scale, the CFA's that are necessary to assess measurement invariance, were ran again. The reason for assessing the modification indices was to signal possible improvements for the scale by rescaling it.

To evaluate the fitness of the models, several absolute, relative and comparative goodness-of-fit indices were used. Goodness-of-fit indices indicate the fit between the statistical model and the observed values (Field, 2000). All these indices are a function of the chi square and degrees of freedom (Hox & Bechger, 1998). First, the absolute goodness-of-fit indices assume that the best fitting model has a fit of zero (Kenny, 2000). The absolute indices that was used in the current study is the Root Mean Square Error of Approximation (RMSEA). The RMSEA value must be below .05 for a good fit and between .05 and 0.8 for a mediocre fit. A RMSEA value above .10 indicates a poor fit. Second, the relative goodness-of-fit indices compare the model to a null model (i.e., the baseline model in which all variables are uncorrelated). The relative indices that were used in the current study are the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI). These values must be higher than .90 for an appropriate fit (Hu & Bentler, 1998). Third, the comparative goodness-of-fit indices are useful when one is comparing multiple models with each other. The comparative indices that were used in this model are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The choice for the indices are based on the checklist for measurement invariance by Van de Schoot, Lugtig and Hox (2012).

The second research question entailed an assessment of the impact of the inability to hold the measurement invariance assumption on the relationship between 'Perceived feedback' and educational outcomes. First, to give an overview of the performance of the students, the mean

and standard deviations of the three proficiency tests (i.e., math, reading and science proficiency) were calculated per country.

Then, the impact of the measurement invariance was assessed with the use of Structural Equation Modeling (SEM), in which the measurement and structural aspect were combined. SEM can be described as a combination of factor analysis and regression analysis or path analysis in which the focus is often on latent (theoretical) constructs (Hox & Bechger, 1998). To answer the second research question, several models were estimated. First, a model was estimated in which the latent variable 'Perceived feedback' is correlated to the scores of one of the proficiency tests. This model was estimated for the whole sample and for the collectivistic countries and individualistic countries separately. Second, a model was estimated wherein the latent variable 'Perceived feedback' is fitted separately for the individualistic and collectivistic countries and again correlated to the scores of one of the proficiency tests. The choice for a correlational relationship between 'Perceived feedback' and the scores of proficiency tests is based on Hattie and Timperley (2007): not only does feedback influence the educational outcomes, feedback also might influence the amount of feedback that the student receives. For an overview of these models, see Figures 2 and 3.

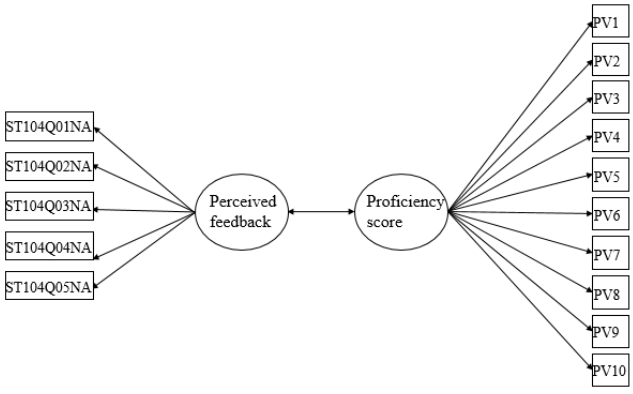

Figure 2. Model 1.

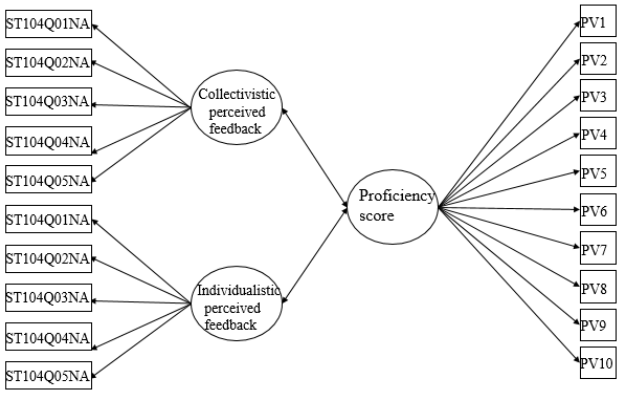

Figure 3. Model 2.

The goodness-of-fit of these models were ascertained and the models are compared with each other. In the case of similar goodness-of-fit indices, the simplest model must be maintained (Hox & Bechger, 1998). Furthermore, the correlations between 'Perceived feedback' and the proficiency score per group of countries are mentioned.

4. Results

4.1. Measurement invariance

First, it was assessed whether the assumption of measurement invariance holds for the 'Perceived feedback'-scale across culturally different nations. To get a first impression of the scale, the mean scores, standard deviation and reliability of the culturally different countries (i.e., collectivistic vs. individualistic) were assessed. To see whether the assumption of measurement invariance could hold for the 'Perceived feedback'-scale, several models needed to be estimated. The analyses started with the most constrained model, because this is the model that is currently used by PISA. Then, the measurement invariance analysis was carried out by using models with different sets of constraints on the factor loadings, intercepts and/or residuals. After this, the model fit was evaluated separately for the groups of collectivistic and individualistic countries and for the country that had the most optimal fit. Finally, the modification indices of the measurement model of the 'Perceived feedback'-scale were evaluated in order to find possible improvements.

Figure 4 displays the distribution of the mean score on the 'Perceived feedback'-scale for the collectivistic and individualistic countries. The collectivistic countries score somewhat higher ($M = 0.278$, $SD = 0.968$) on the 'Perceived feedback'-scale than the individualistic countries ($M = 0.068$, $SD = 0.980$).

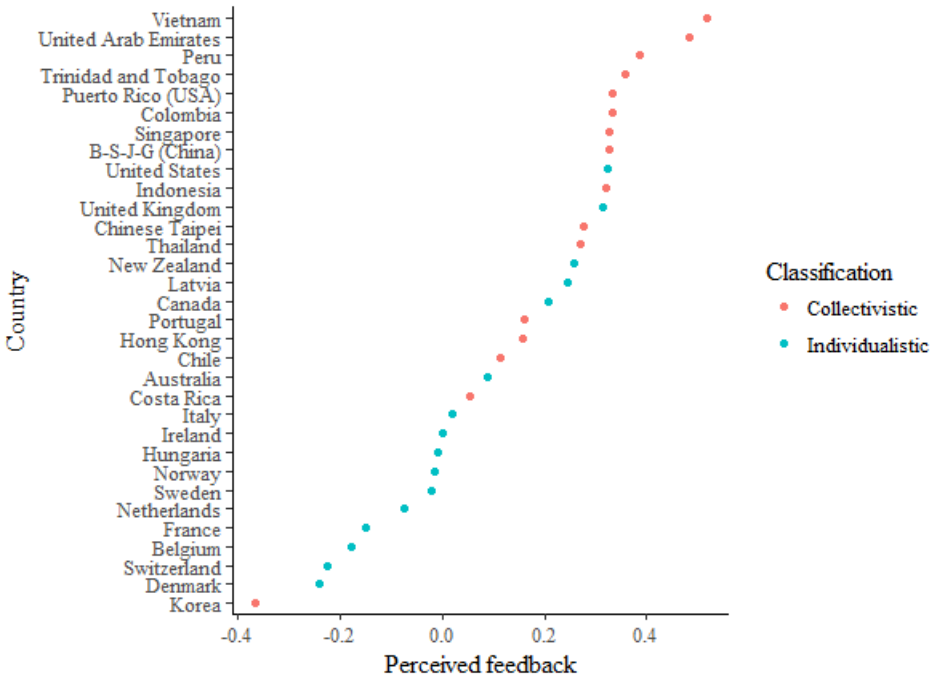

Figure 4. Mean score of 'Perceived feedback'-scale per country.

As can be seen in figure 4, Korea scores much lower (M = -.370, SD = 1.050) than the other collectivistic countries. This shows that the categories are probably not as homogeneous as was expected. Korea might have, for example, educational practices that are more similar to the individualistic countries. For the exact mean and standard deviations of the items and the mean score per country, see Appendix B. In order to assess the reliability of the 'Perceived feedback'-scale, the Cronbach's Alpha of the scale was determined for both the collectivistic and the individualistic countries. For both groups, the reliability of the scale was adequate, namely $\alpha = .91$ for the collectivistic countries and $\alpha = .92$ for the individualistic countries. The reliability has also been determined for each separate country, see Appendix C.

The measurement invariance analysis started with carrying out a CFA on the current model as is used by PISA. This means that this model has constraints on the factor loadings, intercepts, and residuals, and that no distinction is made between the groups. In other words, the same model is used for all countries. The goodness-of-fit indices of the currently used model show mixed results, namely appropriate levels for the relative fit indices (CFI = .966, TLI = .931), but the absolute fit indices is too high (RMSEA = .155). So, the fit of the current model of the 'Perceived feedback'-scale could be questioned and potentially improved. Figure 5 shows the factor loadings. The factor loading of the first question of the scale (ST104Q01NA, "The teacher tells me how I am performing in this course.") is rather low compared to the other questions, which might indicate its unsuitability for the 'Perceived feedback'-scale.

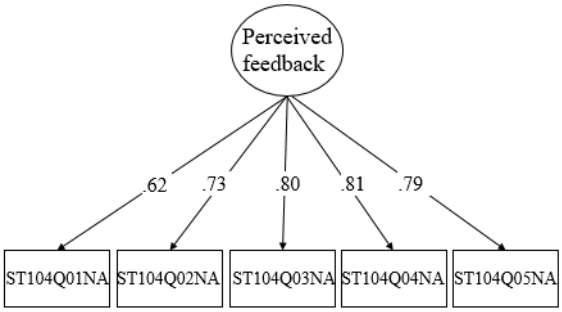

Figure 5. Factor loadings of the model of the current 'Perceived feedback'-scale.

In the next step, the strict measurement invariance analysis was carried out based upon the checklist established by Van der Schoot, Lugtig, and Hox (2012). The equality constraints that are used for these models are increasingly weakened. The goodness-of-fit indices can be found in Table 5. In each model, the relative goodness-of-fit indices, the RMSEA is again found to be too high for an adequate fit. However, the other fit indices are at an appropriate level. This makes the adequacy of the fit of the models somewhat ambiguous.

Table 5. *Results of the measurement invariance analysis.*

Model	χ^2	Df	p	CFI	TLI	RMSEA	BIC	AIC
1 (Strict)	35134	23	.000	.952	.958	.121	2058502	2058327
2 (Scalar)	29665	18	.000	.960	.955	.125	2053094	2052869
3 (Intercept)	29394	14	.000	.960	.943	.142	2052872	2052605
4 (Metric)	25052	14	.000	.966	.951	.131	2048530	2048263

To see whether these models significantly differ from each other, Likelihood Ratio tests were carried out. First, model 1 was compared with model 2 and this resulted in a significant difference, $\Delta\chi^2 = 4342$, $\Delta df = 0$, $p = .000$. Second, model 2 was compared with model 3 and this resulted in significant difference: $\Delta\chi^2 = 271.55$, $\Delta df = 4$, $p = .000$. Third, model 3 was compared with model 4 and this resulted in a significant difference: $\Delta\chi^2 = 5468.3$ with $\Delta df = 5$, $p = .000$. This indicates each model is worse than the previous one and this indicates that the assumption of measurement invariance might not hold for the 'Perceived feedback'-scale across culturally different nations.

To see what the implications are for the baseline models (i.e., the model without any constraints) per country, two CFA's were carried out. The results showed that most goodness-of-fit indices were at an appropriate level, but were worse than the fit of the previous models that were constrained. The RMSEA is again found to be too high for both models, but with a somewhat better fit for the individualistic countries compared to the collectivistic countries. For the values of the goodness-of-fit indices of these models, see Table 6.

Table 6. *Goodness-of-fit indices of the confirmatory factor analysis.*

	χ^2	df	p	CFI	TLI	RMSEA	BIC	AIC
Collectivistic countries	14331	5	.000	.955	.910	.170	1013007	1012864
Individualistic countries	10446	5	.000	.975	.949	.138	1035277	1035133

Figures 6 and 7 show the factor loadings of the baseline CFA models for both the collectivistic countries and the individualistic countries.

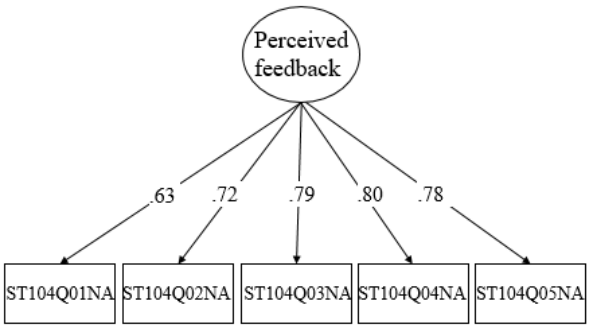

Figure 6. Factor loadings of the collectivistic countries.

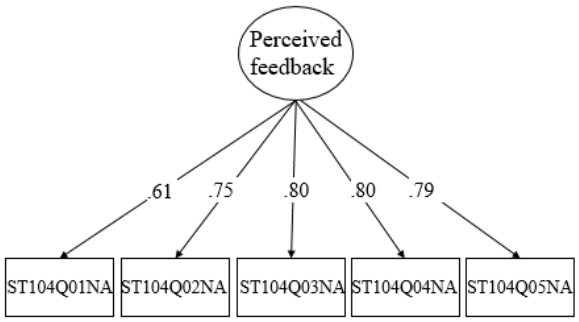

Figure 7. Factor loadings of the individualistic countries.

To compare the model parameters from these models with each other, a CFA with constraints on the factor loadings was compared with the strictest model. The results showed a significant difference, $\Delta\chi^2 = 10082$, $\Delta df = 9$, and $p = .000$. This indicates that the factor loadings are significantly different from each other.

The goodness-of-fit indices of the baseline models show that the individualistic countries have a slightly better fit than the collectivistic countries. Based on this result, the CFA was conducted again for each country separately. The results of these CFA's showed that the model for 'Perceived feedback'-scale generally fitted better for the individualistic countries, see Appendix D. The best fitting model was for The Netherlands. The factor loadings for this country were lower than of the previous models (especially item ST104Q01NA) and are illustrated in Figure 8.

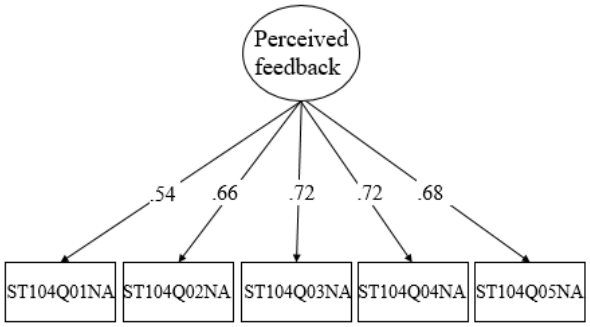

Figure 8. The factor loadings of The Netherlands.

The model still does not have a completely satisfying fit given the fact that the RMSEA is still above the threshold even for the country with the most optimal fit. For this reason, the modification indices were also analysed and these indicated certain adaptions in the model. These adaptions concerned extra correlations between questions ST104Q01NA ("The teacher tells me how I am performing in this course.") and ST104Q02NA ("The teacher gives me feedback on my strengths in this science subject.") and between questions ST104Q04NA

("The teacher tells me how I can improve my performance.") and ST104Q05NA ("The teacher advises me on how to reach my learning goals."). It is very likely that these questions are too similar to each other as the scale is quite narrow and not very specific. In the case of the first two questions, the similarity is that feedback on the current state of performance often includes comments on the strengths of the student. In the case of the last two questions, the similarity is on the Feed Up aspect feedback, namely how to bridge the current state to the desired state. After modifying the model with these adaptations, the measurement invariance analysis was conducted again according to the previous used method. The goodness-of-fit indices of these models were now all of an appropriate level and can be found in Table 7.

Table 7. *Results of the measurement invariance analysis of the adapted model.*

Model	χ^2	df	P	CFI	TLI	RMSEA	BIC	AIC
1 (Strict)	121378	19	.000	.983	.983	.078	2035555	2035339
2 (Scalar)	6950	14	.000	.991	.986	.069	2030427	2030161
3 (Intercept)	6681	10	.000	.991	.982	.080	2030208	2029900
4 (Metric)	2927	10	.000	.996	.992	.053	2026454	2026147

These adaptations in the measurement model of 'Perceived feedback'-scale are also analysed for all countries (i.e., the strictest model). The factor loadings of this model are similar to those of the previous model, see Figure 9. Also, question ST104Q01NA still has a low factor loading.

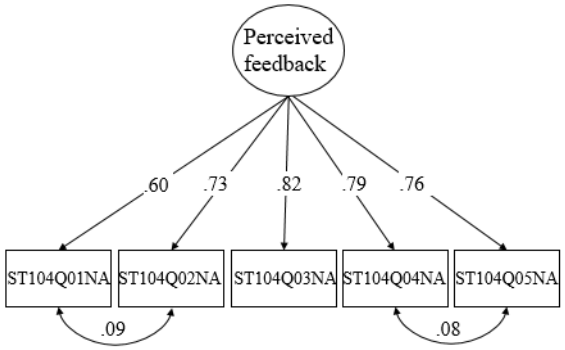

Figure 9. The factor loadings of the adjusted model of the 'Perceived feedback'-scale.

The results of the first research question indicate that assumption of measurement invariance cannot completely be held. The 'Perceived feedback'-scale probably does not completely measure the same perceived feedback across these culturally different countries. The fit of the model is most optimal for the country The Netherlands. By correlating several questions in the model, appropriate levels of model fit can be reached for the 'Perceived feedback'-scale.

However, the first question of the scale (ST104Q01NA) remains to have a low factor loading which indicates its possible inappropriateness in the scale.

4.2. Model fit and educational outcomes

In order to assess the impact of the failure to establish measurement invariance for the 'Perceived feedback'-scale, the impact of culturally different nations on the relationship between 'Perceived feedback' and several educational outcomes (i.e., the math, reading and science proficiency tests of PISA 2015) was addressed. To get a first impression of the math proficiency test scores, reading proficiency test scores, and science proficiency test scores, an overview was given by calculating the mean scores and standard deviations per country. Also, the correlation between 'Perceived feedback' and the educational outcomes was calculated and illustrated with a graph. Then, the impact of the culturally different nations was assessed by estimating several models: one model in which Perceived feedback' was correlated to the score of the proficiency test and another model where this relationship was separated for the collectivistic and individualistic countries. The first model was estimated both for all the countries and for the collectivistic and individualistic countries separately. Also, the models were estimated with in the previous paragraph mentioned modification indices. The significance of the difference between the models determines the impact of the culturally different countries and the need for possible improvements of the model..

Table 8 gives an impression of the math, reading, and science proficiency test scores per collectivistic and individualistic country. The countries are ranked according to the mean score on the math proficiency test.

Table 8. *Mean and standard deviation of the proficiency scores per country.*

	Math		Reading		Science	
	M	SD	M	SD	M	SD
Collectivistic countries						
Singapore	564.19	95.42	405.26	96.63	555.57	103.60
Hong Kong	547.93	90.14	433.54	105.73	523.28	80.57
Chinese Taipei	542.32	102.94	502.90	102.69	532.35	99.55
B-S-J-G (China)	531.3	106.05	484.87	101.14	517.78	103.39
Korea	524.11	99.72	498.52	100.22	515.81	95.18
Vietnam	494.52	83.73	431.72	114.60	524.64	76.60
Portugal	491.63	95.74	407.35	100.19	501.10	91.82
United Arab Emirates	427.48	96.53	526.67	92.79	436.73	99.14
Chile	422.67	85.44	492.20	97.87	446.96	86.02
Trinidad and Tobago	417.24	95.98	458.57	88.14	424.59	93.80
Thailand	415.46	81.50	424.91	89.85	421.34	78.49
Costa Rica	400.25	68.44	427.49	79.24	419.61	70.02
Colombia	389.64	77.17	487.25	100.46	415.73	80.37

Peru	386.56	82.60	509.10	100.12	396.68	76.70
Indonesia	386.11	79.82	499.81	87.30	403.1	68.37
Puerto Rico	378.35	76.86	357.74	84.89	402.65	86.22
Individualistic countries						
Switzerland	521.25	95.75	410.00	95.70	505.51	99.52
Canada	515.65	87.74	499.92	95.68	527.70	92.37
Netherlands	512.25	91.52	433.62	95.06	508.57	100.94
Denmark	511.09	80.57	494.63	87.41	501.94	90.29
Belgium	506.98	97.36	535.10	98.72	502.00	100.19
Ireland	503.72	79.77	452.51	104.23	502.58	88.90
Norway	501.73	84.90	505.22	91.76	498.48	96.24
New Zealand	495.22	92.12	500.16	101.78	513.30	104.09
Sweden	493.92	90.07	497.10	93.20	493.42	102.48
Australia	493.90	93.06	409.13	79.83	509.99	102.30
France	492.92	95.16	427.27	104.07	494.98	101.96
United Kingdom	492.48	92.56	361.06	81.59	509.22	99.65
Italy	489.73	93.57	428.34	82.40	480.55	91.44
Latvia	482.31	77.54	436.57	96.59	490.23	82.21
Hungary	476.83	93.80	496.94	99.85	476.75	96.34
United States	469.63	88.46	486.77	72.60	496.24	98.62

As can be noted from Table 8, the mean scores for the collectivistic countries show a greater difference in scores than the individualistic countries. For example, the math proficiency score of the collectivistic students (range: 378.35 – 564.19) compared to that of the individualistic students. Furthermore, the correlations between ‘Perceived feedback’ and the proficiency scores were assessed. For math and reading proficiency, these were slightly negative, namely -.14 for math proficiency and -.18 for reading proficiency. Due to under-identification of the model (i.e., the model parameters could not be estimated), the correlation for science proficiency could not be calculated. These negative relationships are illustrated separately for the collectivistic and individualistic countries, see Figures 10, 11 and 12.

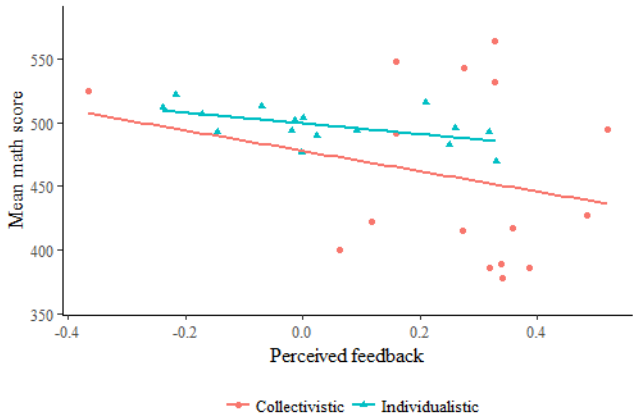

Figure 10. Scatterplot of the relationship between 'Perceived feedback' and the mean math proficiency score.

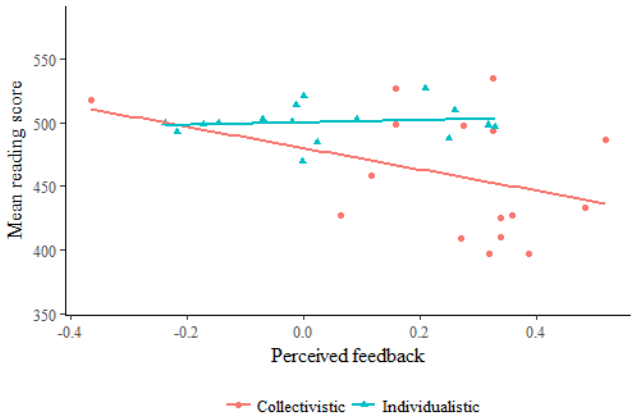

Figure 11. Scatterplot of the relationship between 'Perceived feedback' and the mean reading proficiency score.

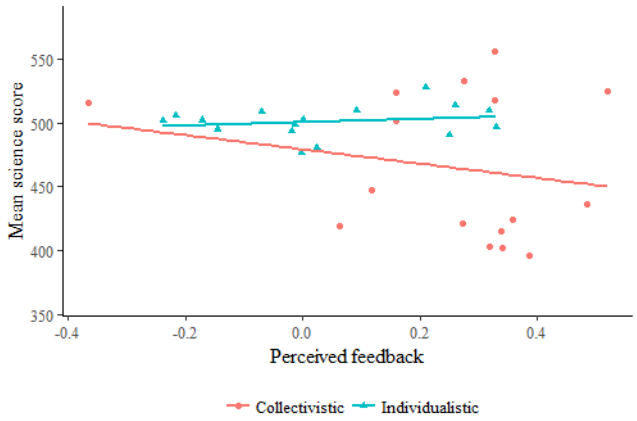

Figure 12. Scatterplot of the relationship between 'Perceived feedback' and the mean science proficiency score.

Figures 10, 11 and 12 show that these relationships differ between collectivistic and individualistic countries. The relationship between 'Perceived feedback' and reading and science proficiency is slightly positive for the individualistic countries and vice versa for the collectivistic countries. Also, from these illustrations it appears that the collectivistic countries are less homogeneous as was expected. The difference in direction between collectivistic and individualistic countries are in line with the fact that measurement invariance cannot be held across these culturally different countries. For this reason, the impact of culturally different nations on the relationship between 'Perceived feedback' and educational outcomes was assessed. Various SEM models, as graphically described in Figures 2 and 3 in the method section, were analysed and compared. First, the impact of measurement invariance was assessed by fitting two models wherein 'Perceived feedback' is

correlated to the proficiency score. The first model estimates the latent variable 'Perceived feedback' for all of the countries included in this study. The second model estimates the relationship separately for the collectivist and individualistic countries. For the goodness-of-fit indices for both models, see Table 9. Second, these models were compared as to see whether the impact of culturally different nations on the relationship between 'Perceived feedback' and proficiency test is significant.

Table 9. *The goodness-of-fit indices for the SEM analyses of the relationships of 'Perceived feedback' with the proficiency tests.*

	χ^2	df	p	CFI	TLI	RMSEA	BIC	AIC
Math proficiency								
Model 1	26190	89	.000	.994	.993	.037	23697563	23697245
Model 2	25788	178	.000	.994	.993	.037	23667598	23666655
Reading proficiency								
Model 1	26714	89	.000	.994	.993	.038	23727742	23727424
Model 2	25969	178	.000	.994	.993	.037	23703025	23702082
Science proficiency								
Model 1	159215	89	.000	.969	.963	.092	22988905	22988587
Model 2	25688	178	.000	.995	.994	.037	22835561	22834618

As can be seen in Table 9, the goodness-of-fit indices showed no great differences between the models. Furthermore, model 1 was compared with model 2 using Likelihood Ratio Tests. This resulted in a significant difference between all models, namely $\Delta\chi^2 = 30712$, $\Delta df = 61$, $p = .000$ for math proficiency, $\Delta\chi^2 = 25464$, $\Delta df = 61$, $p = .000$ for reading proficiency, and $\Delta\chi^2 = 154091$, $\Delta df = 61$, $p = .000$ for science proficiency. As expected, all relationships in the models were found to be slightly negative ($p < .000$). As expected, all relationships in the models were found to be slightly negative ($p < .000$). Table 10 shows the correlations between 'Perceived feedback' and the proficiency tests separately for the collectivistic and individualistic countries. As mentioned, the correlation within model 1 could not be estimated for science when no distinction between the groups of countries is made.

Table 10. *The correlation coefficients between 'Perceived feedback' and the proficiency tests for the various models.*

		Math proficiency	Reading proficiency	Science proficiency
Model 1	All countries	-.14	-.18	-.NA
	Collectivistic countries	-.10	-.14	-.10
	Individualistic countries	-.15	-.17	-.13
Model 2	Collectivistic countries	-.10	-.14	-.10
	Individualistic countries	-.15	-.17	-.13

After adjusting the model of the 'Perceived feedback'-scale according to modification indices (i.e., extra correlations between the first and second question and between the fourth and fifth question), the SEM analyses were conducted again. The SEM analyses did not show any differences between the adjusted model and the prior model, see Table 11. However, the under-identified model of science could be estimated and resulted in a significant correlation of -.13.

Table 11. *The correlation coefficients between 'Perceived feedback' and the proficiency tests for the modified models.*

		Math proficiency	Reading proficiency	Science proficiency
Modified model 1	All countries	-.14	-.18	-.13
	Collectivistic countries	-.10	-.14	-.10
	Individualistic countries	-.15	-.17	-.13
Modified model 2	Collectivistic countries	-.10	-.14	-.10
	Individualistic countries	-.15	-.17	-.13

The second research question addressed the impact of culturally different nations on the relationship between 'Perceived feedback' and educational outcomes. The correlations showed no great differences between the collectivistic and individualistic countries. Furthermore, the goodness-of-fit indices of all models were all of an appropriate level in order to be used. For this reason, it can be assumed that the cultural impact on the relationship between the current 'Perceived feedback' and educational outcomes is negligible and that the first model is appropriate for analysis of 'Perceived feedback'.

5. Conclusions and discussion

The results of the analysis of measurement invariance of the 'Perceived feedback'-scale showed indications that measurement invariance does not completely hold across culturally different countries. One of the fit measures, the RMSEA, was above the allowed value repeatedly, whereas the other fit measures were on an appropriate level. Also, the models were consistently significantly different from each other. This can mean that the latent construct 'Perceived feedback' as measured in PISA has a somewhat different meaning across culturally different nations. Moreover, results showed that the fit of the models for the individualistic countries were to some extent better than those of the collectivistic countries.

The distribution of the mean scores on 'Perceived feedback'-scale showed that individualistic countries perceive somewhat less feedback than collectivistic countries. The difference between perceived feedback and given feedback might play a role here. The students in the collectivistic country might be more inclined to notice important feedback than the students in the individualistic country. The reason for this is that students in an individualistic country are generally more stimulated to be independent thinkers (Staub & Stern, 2002), and might not always easily accept feedback (Hattie & Timperley, 2007). In individualistic countries, it is more common to already have a more student-centred approach (Lau & Lam, 2017). It is likely that an individualistic teacher might have already adapted the lesson to the need of the students, and that feedback is therefore not perceived as such (Hattie & Timperley, 2007). Also, the teacher in the collectivistic country has a greater power distance to the student, which might cause for more notice when feedback is given.

The strictest model (i.e., the model as employed by the OECD in PISA) of the 'Perceived feedback'-scale was estimated and results showed that this model was only mediocre: the relative fit measures were at an appropriate level, but the absolute fit measure RMSEA was well above the acceptable value. This means that the validity of the 'Perceived feedback'-scale currently is not fully satisfactory. This is probably due to the incompleteness of the scale which is also mentioned by Bayer, Klieme and Jude (2016). Further weaker models were estimated to further investigate measurement invariance. The measurement invariance of the models across collectivistic and individualistic nations was estimated on four levels: equality of factor loadings (metric invariance), intercepts, factor loadings and intercepts (scalar invariance), and factor loadings, intercepts and residuals (strict measurement invariance). The results showed that none of the models had an adequate fit due to the RMSEA that again was considered to be too high. This was also the case for the

baseline models that were then estimated for the collectivistic and individualistic countries separately: the RMSEA was slightly too high for the individualistic countries (.138) and much too high for the collectivistic countries (.170). The current model of the 'Perceived feedback'-scale fits best for individualistic country The Netherlands. This is evidence that, indeed, the meaning of what feedback is, somewhat differs across nations that are culturally different with regard to collectivism and individualism. Also, this can mean that the scale as established to some extent better grasps what perceived feedback is in individualistic countries than in collectivist countries. An explanation for this is, as Du Luque and Sommer (2000) mention, that implicit and group-related feedback is the norm in collectivistic countries. The 'Perceived feedback'-scale as implemented in PISA 2015, however, is more focused on the individualistic classroom: (1), the items are all focused on the relationship between the teacher and the student, and (2), all the feedback is directed at the first person, the 'I'. Furthermore, the items specifically focus on feedback that is given *during* the lesson, which might neglect the importance of the feedback given *after* the lesson. Lau and Lam (2017) mention that Chinese (one of the collectivistic countries) teachers are used to focus on the whole class during the lesson and tend to personal issues after the lesson.

Modifying the measurement model of the 'Perceived feedback'-scale based on the modification indices lead to a better model fit. These indices suggested a correlation between the first two questions of the scale and the last two questions of the scale. This finding also indicates that the scale not only lacks some specificity, but that it is also quite narrow. The indices show that the items are very similar to each other: some types of feedback are measured with more than one item, whereas some levels of feedback are not measured at all.

In interpreting the results of this study, it is important to take note of some debate about the individualism-collectivism dimension. Some researchers (Oyserman, Coon, & Kemmelmeier, 2002; Voronov & Singer, 2002) question this categorisation and mention that it is too blunt and does not capture the nuances within the collectivistic and individualistic culture. Evidence for this is found in the 'Perceived feedback'-score of Korea, which does not fit the pattern of the other collectivistic countries. According to Gouveia and Ros (2000), wealth leads to individualism and vice versa. On the other hand, Hofstede et al. (2010) are convinced of its validity and the categorisation has helped cultural psychology to advance (Fiske, 2002). Also, the choice of goodness-of-fit indices for the interpretation of the models is subjective and therefore, some caution in the interpretation of the results of the models is advised.

After assessing measurement invariance, the impact of its mixed results is assessed on the relationship between 'Perceived feedback' and educational outcomes. An impression of the mean scores of the proficiency tests (i.e., math, science, and reading) showed more diverse scores for the collectivistic countries than for the individualistic countries. One explanation could be that collectivistic countries are much more dissimilar in their Social Economic Status (SES) (Gouveia & Ros, 2000). The SES usually impacts the degree of refinement of their educational practices.

The impact of the failure to hold measurement invariance for the 'Perceived feedback'-scale was measured by estimating several models in which the relationship was calculated separately for culturally different countries. First, the relationship was calculated in a model that does not specify the groups (i.e., the collectivistic and individualistic countries). The relationships were, fitting with earlier research, slightly negative for each learning outcome. Second, the model was estimated again, but now separately for the groups. The differences between the relationships between the collectivistic and individualistic countries were only slight. Therefore, it can be assumed that the impact of the failure to hold measurement invariance on the relationship between 'Perceived feedback' and educational outcomes is negligible. Multiple sources (Lau & Lam, 2017; OECD, 2017) explain the negative relationship between 'Perceived feedback' and the educational outcomes by pointing out that worse scoring students might get more attention from the teacher than students that are already perform well. Another explanation is that the given feedback may be not perceived as such. As Hattie and Timperley (2007) mentioned, if the feedback does not fit with the receiver, its power can be weakened or even detrimental. Shute (2007) further elaborates on the specific situations in which feedback is beneficial and made a list of approximately 30 instructions which indicate the details for each appropriate type and level of feedback for each type of student and situation. For example, if the timing of the feedback is unfortunate, this might cause the student to reject the feedback. The items of the 'Perceived feedback'-scale might not be specific enough to assess all these different types and levels of feedback which is in line with the findings of the modification indices and the low factor loadings. The items of the 'Perceived feedback'-scale are mainly focused on the task level: they include questions about the particular science subject, course or learning goals. Feedback focused on the task is often diluted, because it often comes with feedback on the self (Hattie & Timperley, 2007). Moreover, the items are not specific enough whether they refer to knowledge of results (i.e., a grade), knowledge of correct results, or elaborate feedback (Nyquist, 2003). If a student receives feedback on the self or only receives a grade,

it often leads to a diminished effort. The current 'Perceived feedback'-scale might therefore lead one to think that feedback on the task results in a negative learning outcome.

The conclusion of this study is that the current 'Perceived feedback'-scale is feasible to use across culturally different nations, although the fit statistics did partially suggest otherwise. The failure to hold the assumption of measurement invariance does make a negligible impact on the relationship between 'Perceived feedback' and educational outcomes. The same applies for the adaptations of the measurement model of 'Perceived feedback' based on the modification indices. Nevertheless, it would be advisable to replace some questions of the scale as the modification indices pointed out that some were highly similar to each other. These questions could be replaced with questions that assess broader aspects of feedback, such as the cultural aspects. For instance, questions that focus more on the group relation for collectivistic countries and personal relation for individualistic countries can be added. Moreover, peer feedback is more usual in collectivistic countries than in individualistic countries. For this reason, it is very possible that a question that focuses on feedback received from the peer instead of the teacher will be better suited to measure feedback in a collectivistic nation. It is important to note that a potential drawback of replacing items is that the comparability of the scale will be reduced. However, improvement of the scale is regarded as essential in order to adequately measure feedback in the classrooms of both collectivistic and individualistic countries.

For further research, more research into measuring feedback and its different types and levels in the classroom and across nations is recommended. For example, other datasets can be analysed, such as TIMSS context questionnaire. This fits with the suggestion of Bayer, Klieme and Jude (2016), who also mention that further investigation towards the levels and types of feedback across nations would be useful. A further possible recommendation is to consider given feedback instead of perceived feedback, as this can be more objectively ascertained. This way, it would also function as feedback to the teacher self, as it would provoke reflection in the teacher (Hattie & Timperley, 2007).

6. References

Bai, L. M. (2010). Human capital or humane talent? Rethinking the nature of education in China from a comparative historical perspective. Frontiers of Education in China, 5(1), 104–129. doi:10.1007/s11516-010-0008-z

Bayer S., Klieme E., Jude N. (2016) Assessment and Evaluation in Educational Contexts. In: Kuger S., Klieme E., Jude N., Kaplan D. (eds) *Assessing Contexts of Learning. Methodology of Educational Measurement and Assessment*. Springer, Cham. doi: 10.1007/978-3-319-45357-6_19

Black, P., & Wiliam, D. (2004). The formative purpose: Assessment must first promote learning. *Yearbook of the National Society for the Study of Education*, *103*(2), 20-50. doi: 10.1111/j.1744-7984.2004.tb00047.x

Brutus, S., & Greguras, G. J. (2008). Self-Construals, Motivation, and Feedback-Seeking Behaviors. *International Journal of Selection and Assessment*, *16*(3), 282-291. doi: 10.1108/dlo.2009.08123aad.004

Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest and performance. *Journal of Educational Psychology*, *79*, 474–482. doi: 10.1037/0022 0663.79.4.474

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications and programming.* New York, NY: Taylor & Francis.

Byrne, B., & Stewart. S. (2006). The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling, 13*, 287–321. doi: 10.1207/s15328007sem1302_7

Caro, D., & Biecek, P. (2017). intsvy: An R Package for Analyzing International Large-Scale Assessment Data. *Journal of Statistical Software, 81*(7), 1 - 44. doi: 10.18637/jss.v081.i07

Cothran, D. J., Kulinna, P. H., Banville, D., Choi, E., Amade-Escot, C., MacPhail, A.,… Kirk, D. (2005). A cross-cultural investigation of the use of teaching styles. *Research Quarterly for Exercise and Sport*, *76*(2), 193-201.

De Luque, M. F. S., & Sommer, S. M. (2000). The impact of culture on feedback-seeking behavior: An integrated model and propositions. *Academy of Management Review, 25*(4), 829-849. doi: 10.2307/259209

Faitar, G. M. (2006). Individualism versus collectivism in Schools. *College Quarterly*, *9*(4). Retrieved from http://collegequarterly.ca/2006-vol09-num04-fall/faitar.html

Gallagher, M. W., & Brown, T. A. (2013). Introduction to confirmatory factor analysis and structural equation modeling. In *Handbook of quantitative methods for educational research* (pp. 289-314). Springer, Rotterdam. doi: 10.1007/978-94-6209-404-8_14

Gouveia, V. V., & Ros, M. (2000). Hofstede and Schwartz s models for classifying individualism at the cultural level: their relation to macro-social and macro-economic variables. *Psicothema*, *12*(Su1).

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

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112. doi: 10.3102/003465430298487

Henderlong Corpus, J., & Lepper, M. R. (2007). The effects of person versus performance praise on children's motivation: Gender and age as moderating factors. *Educational psychology*, *27*(4), 487-508. doi: 10.1080/01443410601159852

Hirschfeld, Gerrit & von Brachel, Ruth (2014). Improving Multiple-Group confirmatory factor analysis in R – A tutorial in measurement invariance with continuous and ordinal indicators. *Practical Assessment, Research & Evaluation*, 19(7). Retrieved from http://pareonline.net/getvn.asp?v=19&n=7

Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online readings in psychology and culture*, *2*(1), 8. doi: 10.9707/2307-0919.1014

Hofstede, G. (1986). Cultural differences in teaching and learning. *International Journal of intercultural relations*, *10*(3), 301-320. doi: 10.1016/0147-1767(86)90015-5

Hofstede, G., Hofstede G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind*. Revised and Expanded 3rd Edition. New York: McGraw-Hill.

Hox, J. J., & Bechger, T. M. (1998). An introduction to structural equation modeling. Retrieved from http://joophox.net/publist/semfamre.pdf

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological methods*, *3*(4), 424. doi : 10.1037/1082-989X.3.4.424

Hwang, A., & Francesco, A. M. (2010). The influence of individualism–collectivism and power distance on use of feedback channels and consequences for learning. Academy *of Management Learning & Education*, *9*(2), 243-257. doi: 10.1177/0022022102239156

Kaplan, D., & Kuger, S. (2016). The methodology of PISA: Past, present, and future. In *Assessing Contexts of Learning* (pp. 53-73). Springer, Cham. doi: 10.1007/978-3 319-45357-6_3

Kaur, A., & Noman, M. (2015). Exploring Classroom Practices in Collectivist Cultures Through the Lens of Hofstede's Model. *The Qualitative Report*, *20*(11), 1794. Retrieved from https://nsuworks.nova.edu/tqr/vol20/iss11/7

Kenny, D. A. (2000). *Measuring Model Fit.* Retrieved from http://davidakenny.net/cm/fit.htm

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological bulletin*, *119*(2), 254. doi: 10.1037/0033-2909.119.2.254

Kung, M.C. (2008). *Why and how do people seek success and failure feedback? A closer look at motives, methods and cultural differences.* Unpublished doctoral dissertation, Florida Institute of Technology.

Lau, K. C., & Lam, T. Y. P. (2017). Instructional practices and science performance of 10 top-performing regions in PISA 2015. *International Journal of Science Education*, *39*(15), 2128-2149. doi: 10.1080/09500693.2017.1387947

McMillan, J.H. (ed.) (2007), Formative Classroom Assessment: Theory into Practice, Teachers College Columbia University, New York.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525-543. doi: 10.1007/BF02294825

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in crosscultural research. *International Journal of psychological research*, *3*(1). Retrieved from http://revistas.usbbog.edu.co/index.php/IJPR/article/view/857

Nyquist, J. B. (2003). The benefits of reconstruing feedback as a larger system of formative assessment: A meta-analysis. Unpublished Master of Science thesis, Vanderbilt University.

OECD (2016). *PISA 2015 Results (Volume II): Policies and Practices for Successful Schools*, OECD Publishing, Paris. doi: 10.1787/9789264267510-en

OECD (2017). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Proficiency and Collaborative Problem Solving,* OECD Publishing, Paris. doi: 10.1787/9789264281820-en

Prosser, M., & Trigwell, K. (1998). *Teaching for learning in higher education.* Milton Keynes, United Kingdom: Open University Press.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioural Science, 28*(1), 4–13. doi: 10.1002/bs.3830280103

R Development Core Team (2015). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Retrieved from http://www.R-project.org.

RStudio Team (2015). *RStudio: Integrated Development for R.* RStudio, Inc., Boston, MA. Retrieved from http://www.rstudio.com/.

Rosseel, Y., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., ... & Chow, M. (2017). *Package 'lavaan'.* Retrieved from https://cran.r-project.org/web/packages/lavaan/lavaan.pdf

Hothorn (2018). *Package 'lmtest'*. Retrieved from https://cran.r-project.org/web/packages/lmtest/lmtest.pdf

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*(4), 210-222. doi: 10.1002/bs.3830280103

Shepard, L. A. (2006). Classroom assessment. In Brennan, R. L. (Ed.), Educational measurement (4th ed., pp. 623–646). Westport, CT: Praeger.

Shute, V. J. (2008). Focus on formative feedback. *Review of educational research, 78*(1), 153-189. doi: 10.3102/0034654307313795

Staub, F. C., & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains: Quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology, 94,* 344-355. doi: 10.1037/0022-0663.94.2.344

Steenkamp, J. B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross national consumer research. *Journal of consumer research*, *25*(1), 78-90. doi: 10.1086/209528

Suhoyo, Y., van Hell, E. A., Prihatiningsih, T. S., Kuks, J. B., & Cohen-Schotanus, J. (2014). Exploring cultural differences in feedback processes and perceived instructiveness during clerkships: replicating a Dutch study in Indonesia. *Medical teacher, 36(3),* 223-229. doi: 10.3109/0142159X.2013.853117

Van De Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486-492.

doi: 10.3389/fpsyg.2013.00770

Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Measurement invariance. *Frontiers in psychology*, *6*, 1064. doi: 10.3389/fpsyg.2015.01064

Wang, J. (2013). Confucian heritage cultural background (CHCB) as a descriptor for Chinese learners: The legitimacy. Asian Social Science, 9(10), 105–113. doi: 10.5539/ass.v9n10p105

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4. Retrieved from http://ggplot2.org.

Appendix A. The PISA countries and their place on the Individualism Index (Hofstede, 1986).

Score rank	*Country or region*	*IDV score*	*Score rank*	*Country or region*	*IDV score*
1	United States	91	50/54	Bulgaria	30
2	Australia	90	50/54	Dominican Republic	30
3	United Kingdom	89	50/54	Jordan	30
4/6	Canada	80	50/54	Mexico	30
4/6	Hungary	80	50/54	Romania	30
4/6	Netherlands	80	55/57	Portugal	27
7	New Zealand	79	55/57	Puerto Rico (USA)	27
8	Italy	76	55/57	Slovenia	27
9	Belgium	75	58/59	Hong Kong	25
10	Denmark	74	58/59	United Arab Emirates	25
11/12	France	71	60	Chile	23
11/12	Sweden	71	61/65	Albania	20
13/14	Ireland	70	61/65	Singapore	20
13/14	Latvia	70	61/65	Vietnam	20
15	Norway	69	61/65	Thailand	20
16	Switzerland	68	61/65	B-S-J-G (China)	20
17	Germany	67	66	Korea	18
18	Finland	63	67	Chinese Taipei	17
19/23	Estland	60	68/69	Peru	16
19/23	Iceland	60	70	Trinidad and Tobago	16
19/23	Lithuania	60	71	Costa Rica	15
19/23	Luxembourg	60	72	Indonesia	14
19/23	Poland	60	73	Colombia	13
24	Malta	59		Algeria	NA
25	Czech Republic	58		Georgia	NA
26	Austria	55		Kosovo	NA
27	Israel	54		Macao	NA
28	Slovak Republic	52		Moldova	NA
29/30	Spain	51		Montenegro	NA
29/30	Spain (Regions)	51		Qatar	NA
31/32	Japan	46		Tunesia	NA
31/32	Argentina	46		Fyrom	NA
33	Lebanon	40		Massachusettes (USA)	NA
44	Russian Federation	39		North Carolina (USA)	NA
45	Brazil	38			
46	Turkey	37			
47	Uruquay	36			
48	Greece	35			
49	Croatia	33			

Note. When PISA countries have the same ranking, this is indicated with the ranking range. When PISA countries were not included in the study by Hofstede et al. (2011), this is indicated by 'NA'.

Appendix B. The mean and standard deviation of the scores on the 'Perceived feedback'-scale.

	Perceived feedback	ST104Q02 NA	ST104Q02 NA	ST104Q03 NA	ST104Q04 NA	ST104Q05 NA
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Individualistic countries						
U.S.A.	.33 (1.12)	2.39 (0.93)	2.27 (0.99)	2.31 (1.00)	2.36 (0.99)	2.36 (1.02)
U.K.	.32 (0.92)	2.27 (0.79)	2.26 (0.85)	2.38 (0.84)	2.34 (0.85)	2.30 (0.89)
New Zealand	.26 (0.97)	2.23 (0.83)	2.19 (0.88)	2.32 (0.88)	2.33 (0.88)	2.25 (0.92)
Latvia	.25 (0.95)	2.37 (0.90)	2.13 (0.91)	2.26 (0.91)	2.32 (0.92)	2.25 (0.96)
Canada	.21 (1.02)	2.30 (0.84)	2.19 (0.93)	2.22 (0.94)	2.26 (0.94)	2.24 (0.96)
Australia	.09 (0.98)	2.12 (0.82)	2.10 (0.86)	2.17 (0.87)	2.17 (0.88)	2.12 (0.90)
Italy	.02 (0.90)	2.16 (0.82)	1.82 (0.86)	2.10 (0.88)	2.16 (0.89)	2.10 (0.92)
Hungary	.00 (0.92)	2.11 (0.82)	1.94 (0.87)	1.96 (0.87)	2.18 (0.89)	2.07 (0.93)
Ireland	.00 (0.92)	2.05 (0.78)	1.98 (0.84)	2.11 (0.84)	2.09 (0.85)	2.05 (0.89)
Norway	-.01 (0.97)	2.04 (0.83)	1.94 (0.86)	2.09 (0.84)	2.08 (0.85)	2.05 (0.87)
Sweden	-.02 (1.02)	2.05 (0.85)	1.92 (0.90)	2.05 (0.89)	2.10 (0.89)	2.09 (0.91)
Netherlands	-.07 (0.88)	2.03 (0.77)	1.88 (0.81)	2.00 (0.83)	2.04 (0.81)	2.03 (0.83)
France	-.15 (0.96)	1.91 (0.89)	1.72 (0.89)	1.96 (0.90)	2.06 (0.91)	2.00 (0.94)
Belgium	-.17 (0.90)	2.06 (0.85)	1.75 (0.84)	1.90 (0.85)	1.98 (0.85)	1.90 (0.87)
Switzerland	-.23 (1.01)	2.00 (0.88)	1.71 (0.89)	1.87 (0.91)	1.94 (0.92)	1.93 (0.94)
Denmark	-.24 (0.91)	1.85 (0.76)	1.88 (0.82)	1.87 (0.80)	1.88 (0.81)	1.87 (0.83)
Collectivistic countries						
Vietnam	.52 (0.70)	2.36 (0.82)	2.14 (0.84)	2.18 (0.87)	2.78 (0.86)	2.84 (0.87)
U.A.E.	.48 (1.01)	2.44 (0.90)	2.37 (0.94)	2.47 (0.95)	2.49 (0.95)	2.49 (0.98)
Peru	.39 (0.85)	2.26 (0.80)	2.19 (0.85)	2.34 (0.87)	2.44 (0.87)	2.55 (0.94)
Trinidad and Tobago	.36 (1.02)	2.27 (0.91)	2.24 (0.94)	2.38 (0.99)	2.43 (0.99)	2.41 (1.05)
Colombia	.34 (0.94)	2.30 (0.84)	2.16 (0.90)	2.34 (0.91)	2.39 (0.90)	2.42 (0.96)
Puerto Rico	.34 (1.05)	2.36 (0.97)	2.23 (0.96)	2.28 (1.02)	2.41 (1.02)	2.44 (1.05)
Singapore	.33 (0.92)	2.27 (0.78)	2.22 (0.85)	2.34 (0.85)	2.37 (0.86)	2.37 (0.87)
B-S-J-G China	.33 (0.96)	2.08 (0.88)	2.18 (0.91)	2.44 (0.87)	2.33 (0.89)	2.42 (0.90)
Indonesia	.32 (0.80)	1.96 (0.85)	1.94 (0.87)	2.22 (0.91)	2.56 (0.93)	2.65 (0.94)
Chinese Taipei	.28 (0.98)	2.13 (0.84)	2.21 (0.89)	2.32 (0.90)	2.34 (0.91)	2.36 (0.90)
Thailand	.27 (0.83)	2.15 (0.76)	2.22 (0.83)	2.19 (0.82)	2.26 (0.83)	2.46 (0.87)
Portugal	.16 (1.01)	2.12 (0.87)	2.01 (0.90)	2.16 (0.91)	2.30 (0.88)	2.27 (0.93)
Hong Kong	.16 (0.92)	2.16 (0.76)	2.08 (0.83)	2.21 (0.82)	2.24 (0.83)	2.20 (0.84)
Chile	.12 (1.03)	2.14 (0.89)	2.07 (0.93)	2.13 (0.96)	2.23 (0.96)	2.21 (1.00)
Costa Rica	.06 (1.10)	2.16 (1.00)	1.90 (0.97)	2.05 (0.99)	2.19 (0.99)	2.29 (1.04)
Korea	-.37 (1.05)	1.76 (0.86)	1.63 (0.85)	1.72 (0.88)	1.84 (0.91)	1.98 (0.92)

Appendix C. Cronbach's Alpha per country

	Cronbach's Alpha
Individualistic countries	
United States	.944
Sweden	.944
Norway	.940
Australia	.939
New Zealand	.934
United Kingdom	.932
Canada	.931
Ireland	.919
Switzerland	.915
Denmark	.913
Netherlands	.910
Latvia	.898
France	.891
Hungaria	.888
Belgium	.882
Italy	.878
Collectivistic countries	
Korea	.943
Hong Kong	.940
Portugal	.940
Singapore	.933
Chinese Taipei	.932
B-S-J-G (China)	.922
Costa Rica	.921
Chile	.918
United Arab Emirates	.918
Colombia	.906
Trinidad and Tobago	.902
Puerto Rico (USA)	.900
Thailand	.881
Peru	.877
Indonesia	.794
Vietnam	.761

Appendix D. The goodness-of-fit indices per collectivistic and individualistic country.

	CFI	TLI	RMSEA	AIC	BIC
Collectivistic countries					
Portugal	.964	.928	.179	44520.136	44618.043
Puerto Rico	.972	.944	.130	14990.466	15068.283
United Arab Emirates	.966	.932	.157	117307.792	117418.161
Hong Kong	.965	.930	.179	28729.556	28822.871
Chili	.960	.920	.171	60152.951	60253.263
B-S-J-G (China)	.953	.906	.189	84510.912	84617.492
Singapore	.965	.930	.175	46646.767	46746.291
Thailand	.962	.925	.139	69160.806	69262.836
Vietnam	.895	.790	.158	66697.238	66797.183
Korea	.949	.899	.224	41707.192	41805.165
Chinese Taipei	.959	.918	.190	61510.070	61612.911
Peru	.963	.926	.136	58086.591	58186.097
Trinidad and Tobago	.972	.944	.132	45634.111	45729.207
Costa Rica	.975	.950	.135	55221.692	55320.081
Indonesia	.830	.660	.223	73862.315	73963.434
Colombia	.979	.957	.118	105249.541	105358.484
Individualistic countries					
U.S.A.	.981	.961	.138	46204.309	46203019
Australia	.975	.949	.151	91113.908	91223.269
Great Britain	.978	.957	.134	102826.598	102937.690
Canada	.973	.946	.152	158877.491	158993.799
Hungary	.974	.948	.120	43830.770	43926.300
Netherlands	.981	.962	.111	36386.375	36381.156
New-Zealand	.973	.947	.151	33103.202	33196.804
Italy	.972	.944	.122	102998.417	103106.580
Belgium	.961	.923	.146	76605.478	76709.728
Denmark	.958	.917	.171	52738.304	52839.010
France	.964	.929	.144	52738.589	53772.889
Sweden	.967	.935	.177	40498.380	40595.681
Ireland	.965	.929	.161	44805.338	44903.219
Latvia	.965	.930	.145	45709.546	45805.273
Norway	.966	.932	.176	40874.573	40972.286
Switzerland	.973	.945	.139	44702.657	44799.056