7. Discussion

Before discussing the results of this study in detail, we consider it worthwhile to begin by pointing out some general tendencies. The overall impression is that data obtained from parents in both PIRLS and PISA tended to have a higher amount of nonresponse compared to the data provided by students. In some cases, the median nonresponse rate in the parental data was high, at about 15%. In one case (occupation of mother in PISA), the amount of missing data was about 40%. This exceptionally high rate had two origins. First, the nonresponse rate was again about the same size as that for other information collected from parents (less than 20%). Second, about 20% of the parents answered that the mother was working as a housewife and therefore did not have a paid occupation. In this case, missing data could be compensated for by using the information pertaining to both the mother's and the father's occupation as a combined indicator for the family (e. g., the highest ISEI score derived from the occupations coded using the ISCO 88). This way, only 2% of the students were left with missing data for the combined data in addition to the unit nonresponse.

The reasons why information collected from parents consists of a much higher nonresponse rate could be many and different, ranging from lack of interest to unwillingness to complete the questionnaire or separate parts of it. Some parents may not have been willing to provide information they considered too sensitive (private). Given the higher nonresponse rates of parents compared to students, one could argue about whether there really are benefits to including parents as a source of information in large-scale assessment studies. First of all, the extra costs associated with developing, administering, and processing the data from an additional parental questionnaire are considerable. Secondly, research shows that information from Grade 8 and/or 15-year-old students about their families tends to be as valid and reliable as the information provided by their parents (see Chapter 3.5). Nevertheless, because parental occupation has been identified as an important indicator of family background, collecting this data seems a worthwhile exercise. For the older age group in the studies of interest (Grade 8 and/or 15-year-old students), information on parents' occupation could be collected from the students, whereas for the younger age group (Grade 4 students) the preferred option appears to be that of obtaining the information directly from parents. The same recommendation seems applicable to information on parents' education.

The different domains of family background (home possessions, parental education, etc.) showed different strengths of associations with achievement across subject domains and studies. Of all the indicators that we analyzed, the frequency with which the language of the test was spoken at home (the only nondichotomous language variable) showed the weakest relationship with achievement for all subjects amongst all domains.

7.1 Home Background Domains

7.1.1 Immigration Status

The variables related to immigration status generally showed a weak association with achievement. However, in TIMSS Grade 4, the median amount of variance explained by whether or not the students were born in the country of the test was moderate for both mathematics and science. Parents' immigration status seemed to play only a minor role. In regard to the younger group (the Grade 4 students), the information on whether the student was born in the country of the test had more explanatory power than the information on whether his or her parents were born in the country of the test. This pattern was similar for the TIMSS Grade 8 data, but the absolute strength of both associations was much lower than for those emerging from the TIMSS Grade 4 data.

The PISA data showed barely any relationship between the immigration status of students and achievement. For the majority of the countries participating in PISA 2006, the association with achievement was low, although in some countries the regression results were statistically significant, suggesting that immigration status might play a role with respect to the achievement of the younger age group (at least for TIMSS Grade 4 students) but not of the older age group (PISA and TIMSS Grade 8 students). In PISA, the median relationship between the age at which a student immigrated to the country of the test and achievement was negative. Thus, the older students are when they immigrate, the lower their achievement tends to be.

This finding could be simply explained by the fact that students might need to learn the language first before being able to follow instructions in school, with the latter becoming more complex and demanding with higher grade level because of the increasing difficulty of the content taught. However, we cannot regard this finding as a general tendency because, in the majority of countries, the correlation between the variable and achievement in the different subjects varied considerably across the countries and was very weak (below 0.10 in absolute terms). In some countries, it was even positive, but again very weak.

Nevertheless, immigration status is, and likely will remain, a prominent issue on the political agenda, mainly because it is one of the major aspects of equal opportunities in education. International comparative research in education should continue to include the immigration status of students as a characteristic of family background. Even more, this characteristic needs to be more closely integrated into theoretical frameworks as, for example, aspects of social and cultural capital. This approach, in turn, means giving consideration to the role immigration policies in countries might

play in trying to counteract the types of problems students can face when settling in a new country. The possibilities of using immigration status as a component of a scale or an index are also worthy of investigation.

7.1.2 Occupation

In PISA, all occupational variables showed moderate to strong associations with achievement for all three subject domains. The variables on parental occupation as measures of family background seemed to have the largest impact on the educational outcomes. As Hauser (1994) argued, parental occupation is one of the core components of SES (and also of family background) and, compared to other measures, is much more stable over time. As such, it can be used as a long-term indicator of family income. Although the association between the occupational variables and achievement was substantial, the issue of missing data still needs to be considered. In PIRLS, the parent-reported occupational data had an equal amount of missing data for both parents. In PISA, the amount of missing data was at least twice as high for the mother's compared to the father's job, regardless of whether students or their parents provided the information.

The reason seems to originate mainly from question format. The multiple-choice format used in PIRLS has the advantage of respondents only needing to tick an answer. The effort to write in a job title and a job description, as is the case in the PISA questionnaires, is considerably more onerous, which could be one reason for not answering. Also, providing details about an occupation, as in PISA, requires more knowledge compared to only broadly classifying an occupation, as in PIRLS. On the one hand, respondents might be ashamed about their lack of exact knowledge and decide not to provide any information, even though providing just some broad information could still be useful in terms of classifying the occupation. On the other hand, although PIRLS had lower amounts of missing data on occupational variables, the PISA data were much more detailed, and the ISCO classification scheme could be applied with several hundred categories, providing a much finer classification compared to the nominal 10-point scale used in PIRLS. Finally, the ISCO codes can be recoded into the metric scale of the ISEI. In sum, we consider that the several advantages of the open-ended question format for information on occupation outweigh the advantages of a closed question format.

7.1.3 Parental Education

The association of parental education with achievement for the older age group (TIMSS Grade 8 and PISA) was moderate for all tested subjects. The associations were a little higher for TIMSS, but the differences were very small. It can be argued that of all the domains of family background discussed so far, the variables on parental education together with occupation were the strongest predictors of achievement. Since the introduction of the ISCED classification (see Chapter 3.2.10), the administration of parental education in international comparative education research has become a standard procedure. Several studies have published examples showing how adaptation of the abstract ISCED education levels to qualifications specific to national education systems facilitates international comparisons (Foy & Kennedy, 2008b; Foy & Olson,

2009b; OECD, 2009). Parents' education can also serve for studying the reproduction effect proposed by critical theorists; parents' education consequently has also been incorporated into combined measures of family background. Parental education should thus remain a standard characteristic of family background in international large-scale assessments. Still, the international classification scheme used for parental education, that is, the ISCED classification, should undergo constant examination concerning its best possible fit to countries' education systems.

7.1.4 Home Possessions

We operationalized the domain of home possessions in the three large-scale studies of interest by using the highest number of items amongst all five domains of family background. However, only a few of the items showed substantive association with achievement. In all three studies, number of books at home appeared to be the strongest predictor of achievement. It certainly had a strong association with achievement across the different studies and subject areas investigated. This pattern also applied to the association between number of children's books at home and reading achievement in PIRLS.

In general, the existence of educational aids at home showed a low or even almost no association with achievement. The same applied to items of everyday life, such as dishwashers, cars, televisions, and cellphones. The availability of a computer at home showed a moderate amount of explained variance in PISA for all subject areas, but had less explanatory power with respect to the TIMSS Grade 8 data. In both PIRLS and TIMSS Grade 4, the amount of explained variance in achievement on any subject was small. The association between the family having an internet connection and achievement was also low.

These results suggest that computers play a different role for the different age groups. Students from the younger age group might have been using computers at home much less frequently or for different purposes than were the students from the older age group. Use of computers (e.g., as educational aids) thus becomes an additional aspect of computers at home. This explanation would leave the possession of a computer at home as having lesser importance as an indicator of family wealth.

In general, all country-specific (optional) home possession items showed only a very small amount of explained variance in achievement. Their usefulness in international comparisons is questionable not only because of the low association with achievement, but also because each of the countries participating in the IEA and OECD studies was allowed to choose its own set of possession items, thus compromising comparability. In addition, each one of the optional items had its own value for the specific culture, but the extent to which items with different values compare across countries is not clear from the study reports.

7.1.5 Scales

The reliability of the derived scales (PIRLS and PISA) was moderate for four of the scales and rather low for the remaining three. In general, a number of variables showed a strong association with achievement across the three studies and different

age groups, but not all background domains and single items within the domains showed as strong a relationship as commonly described in the literature. The indices in PIRLS showed satisfactory quality overall, although the reliability of the Index of Home Educational Resources was relatively low (less than 0.65). Nevertheless, this variable had a moderate association with reading achievement. In contrast, the Index of Early Home Literacy Activities had moderate reliability, but the association with achievement was quite weak. For the Index of Home Educational Resources in PISA, the median reliability was quite low, although it showed a strong association with achievement. The Index of Family Wealth Possession in PISA had an almost moderate median reliability, but the association between this variable and achievement was the weakest of all the PISA indices. The highest reliabilities and associations with achievement in PISA that we found were with the Home Possessions Scale and the Index of Economic, Social, and Cultural Status.

7.2 Recommendations

Based on the analyses that we conducted, we offer several recommendations regarding the measurement of family background:

1. With respect to all home-possession variables, we recommend including those that show the highest association with achievement in terms of explained variance, namely, number of books in the home, number of children's books in the home, number of student's own books, and access to a computer. This list applies to all three studies—TIMSS, PIRLS, and PISA. Regardless of a strong desire by single countries to include certain home-possession items (and there might be good reasons to include these on a national level), we suggest some standard for a minimum level of discrimination. Usually, items regarded as important will enter the field trial phase, with efforts then made to verify their appropriateness for inclusion in the main survey phase, the outcomes of which are, of course, reported at a later date. A standard for minimum discrimination of families' home-possession items could enhance opportunities to include items that bear the potential to explain variance in student achievement outcomes.
2. Although TIMSS does not collect occupational data, occupational variables have shown their importance and explanatory power in other studies and also in ours. Single items on employment situation in PIRLS showed only a weak association with achievement. Still, constructing a scale together with other occupation measures might yield a better explanation of results. Because collecting information about parental occupation from Grade 4 students does not seem to be reasonable, future TIMSS cycles might consider using a home questionnaire for that grade. This, of course, would introduce extra costs for both international coordination as well as participating countries in terms of developing, administering, and processing an additional questionnaire. A concern might still be the high rate of nonresponse seen in PISA and PIRLS. However, as the results from the analysis reported in this paper have shown, detailed information about occupation can be fruitfully included in models for explaining education outcomes. In TIMSS Grade 8, the information about parental occupation could be collected from students,

given that other studies have already shown that students at this age can provide reliable data about their parents. Thus, information with reasonable explanatory power could be collected without additional questionnaire development-and-administration costs. Only some additional processing would be necessary, mainly in terms of coding occupation information into classification schemes such as the ISCO.

3. Although PIRLS collects parental occupation data, it does so only on a nominal scale level. PIRLS could therefore also profit from choosing to collect more detailed information on parents' occupation. A finer indicator would be desirable so that the effect of family background on achievement could be analyzed in more depth. Using an open format and coding the answers using the ISCO classification scheme seems to us to be a reasonable option that could result in a metric occupation scale, such as the ISEI. An example from Caro (forthcoming) shows that this lack of detailed information is a drawback of PIRLS because it limits the classification of the collected occupational data in other scale types, such as ISEI, even for research purposes.
4. Should the three studies decide to collect more detailed occupation data, we recommend that they include questions about parents' self-employment status and the number of people they supervise in their work. This would allow derivation of EGP classes and, from there, investigation of their association with the assessment outcome variables. The usefulness of these classes was just one example shown by Baumert and Schümer (2001) when they reported the relationship between EGP class and reading achievement in Germany, using data from the PISA 2000 study.
5. PISA derives several scales of aspects of family background, but the reliability tends to be relatively low for the scales on home educational resources, on cultural possessions, and on family wealth. A review of the composition of the scales and further research on optimized operationalizations of aspects of family background might facilitate improvements to these measures of family background. The family wealth scale, in particular, seems to function somewhat differently in different countries. At first glance, the scale seems to relate to the country's economic development, with Western industrialized countries such as Australia, Norway, Slovenia, and Sweden showing no association between family wealth and student achievement, and less-developed countries such as Colombia, Thailand, and Tunisia showing the highest association among the countries in scope. But then the United States also shows up among the countries with the highest association, and Chinese Taipei among the countries with the lowest association. The country-related aspects of family wealth indicators thus need to be investigated further.
6. TIMSS does not derive any scales or indices of family background, although in general these derived variables have shown their predictive power regarding students' achievement. The creation of such indices in future TIMSS cycles might be considered, given they have shown to be highly beneficial for research purposes. Such scales have already been created by researchers working with data from the large-scale assessment studies. For example, May (2002) conducted research which showed that it is possible to create a reliable, valid, and internationally

comparable SES scale for TIMSS and to use it successfully in analyses of study data. Similar findings come from Caro (forthcoming), who constructed a SES scale for PIRLS, using the available relevant data from the study. The derived SES scale had a satisfactory correlation with achievement and satisfactory reliability across countries, although the crossnational validity analysis yielded unsatisfactory results.

Deriving scales from single items offers several advantages. First, reporting can be linked more directly to latent constructs described as important in a study's framework. Second, scales or indices comprised from several items bear the potential of producing a better representation of characteristics of interest with regard to cultural differences among different societies and countries around the world. For example, the cultural value of items is likely to differ a lot between societies when cultural capital is operationalized (maybe amongst other indicators) as home possessions, thereby leaving the value of reporting on single items in doubt. But a combined scale could account for differential value ascriptions of different home possession items in different cultural environments. Of even more pertinence, a combined scale could comprise different (sets of) home-possession items for different cultures. Third, a simple practical reason favors reporting on scales rather than on item level. When deriving valid data for a respondent's scale score, it is usually not necessary to have all items contributing to the scale with valid data. Thus, even when there are (some) missing data in the single items' data, it could be possible to assign valid scale values/scores to more respondents than if any of the single items or a simple combination of the items were used. The latter would necessitate the application of listwise deletion in the analysis.

A further recommendation can be derived from other research that has used recent data from the TIMSS and PIRLS cycles. Both the TIMSS and PIRLS research teams might consider constructing background indices using more complex methods, following examples from the likes of Van Damme, Liu, Vanhee, and Pustjens (2010), who derived a SES index for PIRLS that showed good quality and accounted for a high amount of explained variance in achievement. In her recently published dissertation, Preuschoff (2010) outlined the potential of using Rasch scales for analyzing and reporting TIMSS and the PIRLS background data, in particular effective learning environments connected to the TIMSS 2011 framework (Mullis, Martin, Kennedy, Trong, & Sainsbury, 2009) and PIRLS 2011 framework (Mullis, Martin, Ruddock et al., 2009). She showed that complex scales can be derived with good-quality statistics from the already existing items used in the two studies. In general with Rasch scaling, the respondents can be put on the same metric as the items. Accordingly, "if the content of an item can be paired with its location on the map for the scale, the Rasch scale can be conveniently described in a way that gives meaning to a country's position on the scale" (Preuschoff, 2010, p. 8). Another advantage of analyzing scales instead of single items is attainment of a higher level of measurement. In most cases, an interval scale level can be achieved with scales or indices instead of the ordinal or even nominal scale level of most of the single items. In short, more analytical techniques and computational operations are available to researchers than might appear to be the case.

As we mentioned in Chapter 3.1.4, the SACMEQ approach to comparing findings could be used to improve the understanding of country differences with regard to students' achievement. By comparing absolute values of socioeconomic background across countries (as opposed to relative measures such as percentiles within a country), the relationship between a country's distribution of student achievement and family background differences within the country can be compared with the same relationships in other countries. Future cycles of TIMSS, PIRLS, and PISA could adopt this approach to provide an alternative view of differences in student achievement across countries.

The research by Gershoff, Aber, Raver, and Lennon (2007; see also Chapter 3.2.2 of this paper) suggests a mediating effect of family hardship on the association of income with cognitive skills. If the issues with measuring income can be dealt with to the extent that good-quality data are collected in terms of response rates and validity, it might be worth checking for this type of mediating effect in future large-scale student assessments. However, there might still be issues with measuring family hardships at the international level as, for example, with the definition of poverty, which is a constantly discussed topic on the political agenda and which is frequently changed within countries to suit political purposes.

As we pointed out in Chapter 3.2.8, TIMSS and PIRLS include school-level measures pertaining to the socioeconomic composition of the student body. Information about the proportion of students from economically disadvantaged homes does not provide information about the neighborhood of the student's home, per se, but researchers could think about using it as a proxy for the neighborhood composition of the student's home, given there is only one school that all young people from a certain area attend (because of the lack of alternative schools in the vicinity). Still, this assumption might only hold for rural areas, and, even then, parents in these areas might be able and willing to spend time and money to send their children to a different school further away. It therefore seems desirable to collect information about the neighborhood of the students' home directly by, for example, asking the parents. Of course, the validity of this information needs to be checked, as there might be an issue with respondents giving answers that they see as socially desirable. Some parents, for example, might feel ashamed to reveal that their family is living in a neighborhood with a bad reputation.

We also suggest paying closer attention to religion as a measure of students' family backgrounds. As we mentioned in Chapter 3.2.9, IEA's ICCS survey includes engagement with religion as a part of broader civic engagement in order to investigate the attitudes of students toward the influence of religion in society. It would be interesting to explore if religion is associated not only with students' attitudes but also with students' achievement.

A final suggestion is to further explore practices within families. PIRLS and PISA both include a few questions about activities in the family directed toward improving the child's reading (PIRLS) and science (PISA) skills. Research, we think, should focus more on family practice with respect to supporting students' skills development.

7.3 Further Research Needs

It was not possible for us, in the current project, to cover all issues related to the measurement of home background in large-scale international assessment studies of educational achievement. The main purpose of our study was to explore some basic issues related to family background measures—nonresponse rate, degree of association with student achievement, and reliability of the scales and indices. And even though these basic characteristics address only some of the issues concerning the measurement of the family background of students, they are still ones that merit ongoing consideration.

In this paper, we briefly addressed the possible role of countries' immigration policies with regard to the effect of the immigration status of the family on students' achievement. Further research could take a closer look at the similarities and differences between the immigration policies of the countries participating in international large-scale assessments.

It would also be desirable to investigate further the use of indicators of family background on a higher aggregated level such as neighborhood characteristics. For example, TIMSS 2007 reported on information provided by school principals about the percentages of students in their respective schools who were from economically disadvantaged families and/or had the language of the test as their native language (Mullis et al., 2008). The inclusion of such variables in a more complex model (multilevel) could reveal the contribution that these variables make in terms of explaining differences in student achievement. Multilevel analyses (e. g., multilevel regression analyses or hierarchical linear modeling) could be used to further explore critical factors of the family's environment, on the one hand, and to differentiate the influence of the factors from different levels (e. g., student, school, and country), on the other. Examples of country-level data that are available from diverse sources could include educational expenditure, achievement orientation in general, and the specific subject domain that is the focus of the respective student assessment.

Another direction relative to more advanced methods of exploring the influence of family background on student achievement was recently proposed by Sandoval-Hernández (2012). The author used exploratory structural equation modeling (ESEM) to explore the factor structure of indicators of economic, social, and cultural capital as operationalized in PIRLS 2006 and PISA 2009. ESEM allows for items to load on more than one factor simultaneously. Sandoval-Hernández's results suggest that more sophisticated analyses of this kind (and which were beyond the scope of our paper) could yield more insight into the structure of family background items.

Finally, we draw attention to two additional important issues related to family background that we consider are very much in need of further research-based exploration. They are the crosscultural validity of the family background measures and their coverage in the studies. We therefore end this paper by sketching further research needs with regard to those two aspects.

7.3.1 Crosscultural Validity

Addressing the measurement of family background in international large-scale assessment studies requires analysis of crosscultural validity that is structure-oriented or, in other words, concerned with determining whether the measures (in our case) of family background used in TIMSS 2007, PIRLS 2006, and PISA 2006, or of any other large-scale assessment study measure the same construct across all countries (see Van de Vijver, 2003a).

We mentioned in Chapter 3 that comparing home possessions across countries is not a straightforward process. Home possessions are not the only domain in which crossnational comparisons can be problematic. According to Braun and Mohler (2003), national characteristics, typical social structures, and legal institutions can make the definition of a latent construct quite a difficult task, and the indicators chosen may not be equivalent across countries and cultures. The same is true for the questions in survey instruments. All of this can result in a comparison of incomparable things (Braun & Mohler, 2003).

The measurement of a construct across countries is usually described with either one of two terms—"cross-cultural bias" and "cross-cultural equivalence." These two terms are generally regarded as antonyms. On the one hand, bias refers to the "presence of nuisance factors that challenge the comparability of scores across cultural groups" (Van de Vijver, 2003a, p. 144). The notion of biased scores lays the focus on cultural differences. On the other hand, equivalence refers to the comparability of scores across cultures. Bias and equivalence are not related to the survey instrument itself, but rather to its applications in different countries. For this reason, the presence or absence of bias or equivalence needs to be determined empirically.

Sources of bias can also reside in the constructs themselves, the methods used, and the items by which the constructs are measured; hence, three different types of bias can be distinguished (see Van de Vijver, 2003a, pp. 145–147). They are construct, method, and item bias.

Construct bias is inherent when the measured construct is not identical across cultures. Sources of construct bias could include the following:

- (Partial) differences in the meaning of a construct definition across cultures;
- Different behaviors associated with the construct across cultures;
- Poor selection of the behaviors that are manifesting the construct (i.e., dimensions of the construct measured by the items in the instrument); and
- Incomplete coverage of the relevant aspects of the construct itself.

Sources of method bias typically include (amongst others):

- Samples that cannot be compared with each or one another;
- Ambiguous instructions for the respondents or survey administrators; and
- Differences in familiarity with the stimulus material or response procedures.

Item bias can occur because of:

- Ambiguous items or poor translations;
- Nuisance factors (e. g., an item invoking additional traits or abilities); and
- Culture-based peculiarities (Van de Vijver, 2003a).

However, scores and scales should not depend on variance that is not relevant to the underlying construct (French & Finch, 2006). As Schulz (2006) points out, measurement of family background of students across countries requires crosscountry validation of the underlying construct. A high nonresponse rate also introduces bias. Keeves, Lietz, Gregory, and Darmawan (2006) point out that, in the case of achievement, for example, bias due to nonresponse inflates the mean level of performance and reduces the variance. This kind of bias would give erroneous estimates of the achievement and also reduce the capability of the analysis of variance techniques used in later analyses (Keeves et al., 2006). As Mullis (2002) also points out, the nonparticipation of students introduces bias in the results by increasing or decreasing the performance. Nonresponse and nonparticipation can also be driven by cultural differences and thus need to be investigated with regard to crosscultural validity.

The aforementioned issues pose a problem of marked current relevance for international comparative studies. Measures and indicators need to be comparable across participating countries and/or education systems. When reporting relations or correlations at the international level, researchers need to be sure that indicators are measuring the same things across countries. The validity of scales (indices) obtained by international large-scale assessment studies (such as TIMSS 2007, PIRLS 2006, and PISA 2006) need to be analyzed. Such an analysis would include only those scales that show a promising quality in terms of nonresponse and reliability, as explored in the previous stages of this paper.

Bias or equivalence analysis endeavors to determine the level of comparability of data across cultures. It also verifies whether nuisance factors are distorting the results. The analysis of bias in crosscultural studies attempts to identify whether nuisance factors associated with variation are present. Equivalence analysis is aimed at determining the consequences of bias on crosscultural comparisons.

Different statistical techniques can be used to explore whether the underlying construct measured is the same across cultures. Researchers wanting to determine crosscultural validity could use multigroup confirmatory factor analysis (MGCFA). Confirmatory factor analysis (CFA) is "one of the several statistical techniques that form part of structural equations modeling" (Van de Vijver, 2003b, p. 212). CFA involves decomposing correlations or covariance and then testing to what extent the observed covariance can be reconstructed. This approach is carried out while assuming a specified (in advance) factor constellation. The specific parameters that need to be estimated can be constrained to be equal, after which the factor loadings, factor covariance, and error variances can be examined (Van de Vijver, 2003b). Constraints

for reviewing the invariance of model parameters can be of different types once the invariance of the factor structure and factor loadings have been examined (Schulz, 2006).

Other methods and models that are often applied in crosscultural validity studies use item response theory (IRT) as a basis. IRT has some attractive features, such as the link between the person and item parameters, which produces results that can be empirically examined. Also, once data have been fitted, IRT allows comparisons of item parameters across cultures. This can be done by using statistical tests of differences on these parameters, and those tests can often then be used to identify items that are biased. In addition, IRT can deal with instruments that are not identical in all countries. Item parameters, however, can be compared across countries regardless of the differences in the scores in each one of them, but only if the underlying latent trait (construct) in all countries is the same (Van de Vijver, 2003b).

This brief overview of the two statistical methods shows that a crosscultural validity analysis of all derived scales and indices used as a measure for student family background would be desirable. It also shows multigroup confirmatory factor analysis or multigroup IRT to be methods that are well suited to this process.

7.3.2 Coverage

The concept of family background refers to many aspects that cannot be covered entirely within one study. There are too many questions that theory states to be important, on the one hand, and too little time and too few resources to enable one to attempt to answer all those questions, on the other. Nonetheless, an important part of future research would be to identify the aspects of family background encompassed by the international large-scale education studies for predicting the achievement of the students assessed. Moreover, this analysis should include not only a listing of the aspects that were actually assessed, but also a side-by-side comparison of the studies with regard to the aspects covered and the number and types of variables used. Content analysis of the assessment frameworks and of the instruments that were used to collect contextual data should be the methodological approach for this kind of research. Such an analysis would help to delineate the differences in the approaches that the large-scale studies that we have considered in this paper have used to measure students' family backgrounds.

Content analysis is a method that has long been used in research. According to Krippendorff (2004, p. 18), content analysis is "a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use." Although some authors argue that content analysis is a quantitative method, Krippendorff (2004) distinguishes qualitative and quantitative alternatives of the technique, and gives preference to the qualitative ones—"Reading is a fundamentally qualitative process, even when it results in numerical results" (Krippendorff, 2004, pp. 19–20)—and states that replacing the exact quoting with numerals is just for convenience (Krippendorff, 2004, p. 87). He then advises identifying and counting the frequency of occurrence of each of them.

This is the reason why the boundary between qualitative and quantitative content analyses is blurred (Priest, 2009, p. 40). Quantitative approaches reduce the complexity of information. This process involves isolating single elements of the information, counting up their specific characteristics, and then classifying them against criteria. Qualitative approaches attempt to identify the meaning of the information as a whole. Identifying the meaning by interpretation is the next step (Kelle, Prein, & Bird, 1995, p. 168). While a distinctive trait of qualitative and quantitative content analyses is their orientation to manifest (in quantitative) and latent (in qualitative) content, either kind of approach often means analyzing the traits ascribed to the other. To put this consideration another way, sometimes the qualitative approach deals with the manifest content and the quantitative with the latent. Nevertheless, the qualitative content analysis is concerned more with the latent content and "can [therefore] better take into account subtleties of the structure of arguments and narratives not easily captured by quantitative summaries" (Priest, 2009, p. 108).

Future research intent on measuring family background from a methodological perspective and focusing in particular on the approaches adopted by international large-scale education studies should use a mixture of qualitative and quantitative variants of content analysis, but pay greater heed to the qualitative. The indicators of family background and the variables derived from them (indices, scales) could then be classified according to the aspects of family background they cover in each of the studies. The studies could then be compared and evaluated with regard to the presence or lack of variables and indicators in each of the aspects that comprise the theories of family (home) background.