Fleckenstein, Johanna; Leucht, Michael; Pant, Hans Anand; Köller, Olaf

Proficient beyond borders: assessing non-native speakers in a native speakers' framework

Large-scale assessments in education 4 (2016) 1, S. 1-19

Large-scale **Assessments** in Education

CrossMark

Proficient beyond borders: assessing non-native speakers in a native speakers' framework

Johanna Fleckenstein[1*], Michael Leucht[1], Hans Anand Pant[2] and Olaf Köller[1]

*Correspondence:
fleckenstein@ipn.uni-kiel.de
[1] Leibniz Institute for Science and Mathematics Education, Olshausenstr 62, 24118 Kiel, Germany
Full list of author information is available at the end of the article

Abstract

Background: English language proficiency is considered a basic skill that students from different language backgrounds are expected to master, independent of whether they are native or non-native speakers. Tests that measure language proficiency in non-native speakers are typically linked to the common European framework of reference for languages. Such tests, however, often lack the criteria to define a practically relevant degree of proficiency in English. We approach this deficit by assessing non-native speakers' performance within a native speakers' framework.

Method: Items from two English reading assessments—the Programme for International Student Assessment (PISA) and the National Assessment (NA) for English as a foreign language in Germany—were administered to $N = 427$ German high school students. Student abilities were estimated by drawing plausible values in a two-dimensional Rasch model.

Results: Results show that non-native speakers of English generally underperformed compared to native speakers. However, academic track students in the German school system achieved satisfactory levels of proficiency on the PISA scale. Linking the two scales showed systematic differences in the proficiency level classifications.

Conclusion: The findings contribute to the validation and international localization of NA standards for English as a foreign language. Practical implications are discussed with respect to policy-defined benchmarks for the successful participation in a global English-speaking society.

Keywords: English language proficiency, Non-native speakers, Reading literacy, National assessment, Linking study, Validation

Theoretical Background

The popularity and relevance of the English language has made foreign language education a mainstream endeavor around the globe (Dixon et al. 2012; Hu 2007). English language proficiency is considered a basic skill (Cha and Ham 2008; Graddol 2006) that students from different language backgrounds are expected to demonstrate, regardless of whether they are native speakers or non-native speakers. This holds true for all dimensions of language proficiency, however, the comprehension of written texts is one of the most important academic skills for students to acquire and improve (Edele and Stanat 2016; Urquhart and Weir 2013). Therefore, reading is considered a major prerequisite

Fleckenstein *et al. Large-scale Assess Educ* (2016) 4:19

Page 2 of 19

for educational success. Despite more or less advantageous linguistic environments and learning opportunities, students need a certain degree of English reading proficiency in order to be able to compete in the globalized economy (Grabe and Stoller 2013).

The relevance of English reading proficiency in the ongoing process of globalization seems underrepresented in the international discourse on educational outcomes. The establishment of accountability systems has been a major issue in national educational policy and research for the past decades. However, explicitly testing English as a language of unique global significance is not usually part of international large-scale assessments.

Since 2000 the Programme for International Student Assessment (PISA) has assessed students' English reading literacy in first language English majority countries. English as a Foreign Language (EFL) proficiency of non-native speakers on the other hand is often measured against the CEFR, which plays a central role in language policy and educational standardization in Europe and beyond (e.g., National Educational Standards in Germany; cf. Köller et al. 2010). The CEFR provides descriptors of what foreign language students are able to do in terms of communicative language competence. But what level of English language proficiency is necessary to master life in a globalized world? PISA can give us an idea of what non-native speakers should be able to do in terms of English reading literacy by comparing them to native speakers and by measuring them against standards that were shown to be predictive for educational and vocational success in English-language contexts.

Up to now, there has been little research on performance discrepancies in the language proficiency of non-native speakers, who are learning the language in a predominantly academic setting, with perhaps some exposure to the language through media and information technology, versus native speakers who are exposed to the language inside and outside of the classroom from an early age. The present study is an attempt to overcome this gap using data pertaining to EFL students in Germany. We administered two English language reading tests on EFL students: one that was constructed specifically for the purpose of assessing proficiency of EFL learners (National Assessment in Germany; NA), and one that was originally constructed to assess reading literacy in the testees' first language (PISA). This allows us (a) to compare non-native speakers (German EFL learners) to native speakers (students from Anglophone countries) on the PISA scale, and (b) to compare proficiency level classifications and policy-defined benchmarks by linking the two test scores for each student in a common-person approach.

English as a global language

English is considered the first global language (Crystal 2006; Gil 2011; Romaine 2006; Svartvik and Leech 2006). It has become the leading language of international discourse that is used in a variety of contexts, most evidently in academic and business communities. Crystal (2006) presented the following criteria that a language must meet in order to be considered a global language: (1) it is the native language of the majority of people in some countries, (2) it has been widely adopted as an official language, and (3) it is a priority in foreign language teaching around the world. Conservatively estimated,

Fleckenstein *et al. Large-scale Assess Educ* (2016) 4:19

Page 3 of 19

approximately 329 million people speak English as a native language; however, they are easily outnumbered by second and foreign language speakers (Crystal 2003a, b).

In 19 of the 27 member states of the European Union (EU) English is the most widely used foreign language—this does not include those countries where it is the native language. Moreover, half of the EFL speakers in the EU use English on a regular basis (Eurobarometer 2012). An increasing number of universities in different parts of the world offer study programs with English as the medium of instruction (Ferguson 2007; Foskett 2010; Jenkins 2014; Maringe and Foskett 2010). Five first language English majority countries (US, UK, Australia, Canada, and New Zealand) host almost half the total of foreign students in tertiary education globally (Jenkins 2014).

Advanced levels of reading comprehension facilitate the acquisition of knowledge and new ideas (Chall 1996). Therefore, it has to be considered a major prerequisite for educational success. The global relevance of reading proficiency in English becomes apparent especially in academia as research is increasingly published in international English-language journals and English-medium university programs have become common practice in many non-anglophone countries. Scholars and university students alike are required to comprehend written texts in the English language. But the context, in which reading in English is a relevant skill, is much broader than that. Reading English texts online, communicating with international business partners via e-mail, understanding transnational policies or contracts—these are examples for everyday situations in which sufficient English reading proficiency is a necessity. Thus, the lack of it would certainly impede successful participation in a global society and labor market.

The prominent status of the English language all over the world can have diverse manifestations. The most common classification of World Englishes by US linguist Braj Kachru (1986, 1988, 1992, 2011) visualizes the spread of English around the globe in a model consisting of three concentric circles. These circles represent different ways in which the language is acquired and currently used by the people of a certain country (see Fig. 1). First of all, there is the so-called Inner Circle of countries in which English is the

Fig. 1 Kachru's classification of English as a global language

Fleckenstein *et al. Large-scale Assess Educ* (2016) 4:19

Page 4 of 19
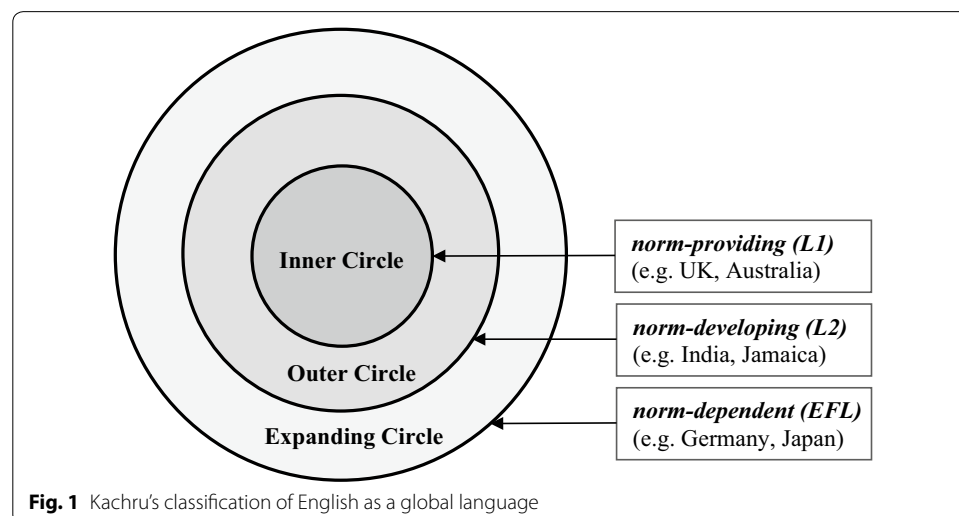
native language of the majority of the people.[1] Kachru called this circle norm-providing since its speakers are usually considered to set the standards for English language proficiency. Secondly, there are the countries of the outer circle that have adopted English as an additional (official) language mostly due to colonization by the British Empire. In the Outer Circle, English is used for intra- and international communication in a multilingual environment. It is the second language for most speakers, acquired rather early and in a natural language environment (e.g., as the language of instruction in school, or in everyday communication between speakers of different languages in one country). Third, there are those countries in which English is learned as a foreign language by almost all people. This is by far the most rapidly growing group, which is why Kachru called it the Expanding Circle. In the Expanding Circle countries English is not necessarily acquired in everyday life but needs to be learned and taught at school. Therefore, students' exposure to authentic communication in English is limited, especially outside of school and particularly concerning spoken and written production ([reference deleted to maintain the integrity of the review process]; for a critical discussion of authenticity see Gilmore 2007). In the Expanding Circle, English is not usually the language of communication inside the borders of a country. It is almost solely used for international communication between people from different native language backgrounds, especially in business and academic settings.

The construct of reading proficiency

Language proficiency is a multidimensional construct that is commonly divided into the four sub skills reading, listening, writing, and speaking. The present study focuses on reading as an indicator of language proficiency, thus, the findings are not necessarily generalizable to the other skills. The rationale for comparing native speakers and non-native speakers on English reading literacy rather than other aspects of language proficiency can be derived from the concept of higher versus basic language cognition: The Common Underlying Proficiency (CUP) model proposed by Cummins (1981, 2000) states that proficiencies involving more cognitively demanding tasks (such as literacy, content learning, abstract thinking and problem-solving) are common across different languages. This assumption can be considered the basis of the linguistic interdependence hypothesis, which posits that certain skills are transferred across different languages in multilingual individuals (Cummins 1979). According to Hulstijn (2011) reading (as well as writing) is an aspect of language proficiency that requires higher language cognition (HLC) as opposed to (just) basic language cognition (BLC). All adult native speakers easily achieve a high level of BLC, however, they are expected to differ in their HLC profiles depending on their intellectual skills, education, professional careers and leisure-time activities. Hulstijn (2011, p. 242) claimed that "while L2 learners can acquire HLC in their L2 as native speakers can, it remains an open question to what extent postpuberty L2 learners can fully acquire BLC in their L2". However, as long as they share a similar intellectual, educational, professional, and cultural profile, second or foreign language learners can reach the same level of HLC as native speakers, despite some deficiencies

---

[1] There are non-native speakers in the Inner Circle as well, namely those who (or whose parents) migrated from an outer or Expanding Circle country. These students are usually referred to as English as an additional language (EAL) students or English language learners (ELLs).

Fleckenstein *et al. Large-scale Assess Educ* (2016) 4:19

Page 5 of 19

in their BLC. Thus, comparing students who differ in their first language but otherwise have similar backgrounds makes more sense for HLC (as in reading literacy) than for BLC (as in speaking). According to Hulstijn's theory, EFL learners would actually be capable of reaching similar levels of reading proficiency as those with English as their first language.

Urquhart and Weir (2013, p. 21) gave a very basic definition of reading as "the process of receiving and interpreting information encoded in language form via the medium of print". Reading involves the reader, the text, and the interaction between the reader and text (Rumelhart 1977; Kintsch and Mangalath 2011). This holds true for reading in a first, second, or foreign language, as many studies have pointed to processing similarities first language (e.g., Droop and Verhoeven 2003; Jongejan et al. 2007; Lesaux et al. 2007). So while first, second, and foreign language reading seem to share many similar features, the processes also differ to a certain extent (Birch 2014; Koda 2005). Enright et al. (2000) indicated three fundamental differences: Second or foreign language readers build on prior first language reading experience, their reading processes are cross-linguistic, involving two or more languages, and their reading instruction usually commences before adequate oral proficiency in the target language has developed. Thus, second or foreign language learners usually have not mastered the basic language structure prior to reading instruction and—compared to first language readers—they are not continuously exposed to written language in their cultural environment (De Zeeuw et al. 2013; Martin et al. 2013). These differences lead to qualitatively different comprehension processes in first, second or foreign language reading. As such, the uniqueness of second and foreign language reading is due to (a) the transfer of first language reading skills and strategies, (b) the facilitation resulting from structural similarity in first language and second or foreign language, (c) the cross-linguistic interactions during second or foreign language reading, and (d) the processing constraints imposed by limited linguistic knowledge. Against the background of these similarities and differences in first language and second or foreign language reading, in the following we address the constructs of reading proficiency as defined by PISA and German NA, how they were operationalized and classified into proficiency levels.

Common assessment frameworks for English reading proficiency

Kachru's three concentric circles represent different ways in which English is used and acquired by the people of a certain country or region. Assessment frameworks for English language proficiency differ accordingly: In the Inner Circle the PISA reading literacy test has been used to assess proficiency in native speakers, while in the Expanding Circle non-native speakers are usually assessed by tests linked to CEFR. This is also the case for German NA, which includes a test for EFL reading comprehension. There are two different models of proficiency level classifications that correspond to these two frameworks. Both consist of a scale with five levels for the localization of student abilities and both of them set certain standards for English language proficiency.

Test of reading literacy in PISA for native speakers

PISA is an international comparative study by the Organization for Economic Co-operation and Development (OECD) of 15-year-old students. The literacy concept in PISA

Fleckenstein *et al. Large-scale Assess Educ* (2016) 4:19

Page 6 of 19

aims to assess skills for life of 15-year-old students, which refers to competencies necessary for participation in society and success on the labor market (OECD 2003). The PISA test for English (i.e., inner circle, native speakers, thus first language) reading literacy measures "an individual's capacity to understand, use and reflect on and engage with written texts, in order to achieve one's goals, to develop one's knowledge and potential and to participate in society" (OECD 2009, p. 14). It examines to what extent adolescents are able to understand and integrate texts they are confronted with in their everyday lives. PISA measures "students' applied ability to deal with written material through handling different kinds of text and performing different types of reading tasks in relation to various situations where reading is needed" (OECD 2004, p. 272). The targeted reading tasks include (1) retrieving information, (2) forming a broad understanding, (3) developing an interpretation, (4) reflecting on the content of a text, and (5) reflecting on the form of a text. The test includes continuous and non-continuous texts.

The construct of reading literacy is inferred from the responses of students on a number of items. Until PISA 2012 a Rasch model was used to draw plausible values (PVs) for student abilities which are then transformed to a metric with the OECD mean score of 500 and a standard deviation of 100. Thus, what a student with a certain score is able to do, i.e., what kind of reading tasks he or she can solve with sufficient certainty, can be described in relation to task demands. Based on these descriptions, five proficiency levels (six from PISA 2009 onwards) are provided, each covering a certain range on the ability scale (see Table 1) and characterized by certain demands of reading tasks.

There are certain prominent benchmarks for reading literacy that are considered relevant for lifelong learning and successful participation in society. The majority (57%) of 15-year-old students in the OECD is proficient at level III or above. It is also the most common level of highest performance for students across OECD countries (OECD 2010b). Thus, in the following we will refer to level III as *norm proficiency*. Reading proficiency at this level has been characterized as the ability to "compare, contrast and categorise competing information according to a range of criteria" (Bussière et al. 2001, p. 24). Level III is often considered a key measure of success in PISA and it is used as a benchmark for national standards in a number of countries (e.g., Canada, Australia). The Canadian Youth In Transition Survey (YITS) could show that the PISA reading literacy

Table 1 Classification of proficiency levels/standards and the respective scoring range for PISA and German NA (CEFR)

PISA		German NA (CEFR)	
Proficiency level	Score on PISA scale	Proficiency level	Score on NA scale
I	≤407	A1	≤399
II	408–479	A2	400–499
III	480–552	B1	500–599
IV	553–625	B2	600–699
V	≥626	C1	≥700
Proficiency standards			
Baseline proficiency (II)	≥408	Minimum standard (A2.2)	≥450
Norm proficiency (III)	≥480	Norm standard (B1.2)	≥550

Fleckenstein *et al. Large-scale Assess Educ (2016) 4:19*

Page 7 of 19

level III is an important level of achievement for predicting post-secondary academic success (Bussière and Knighton 2006) as well as the reading achievement of 21 and 24 year olds, respectively (OECD 2010a, 2012). Likewise, the Swiss youth panel Transitions from Education to Employment (TREE) found young people with poor reading skills (level II or below) are three times more likely to drop out of post-compulsory education (24% dropout rate) than those commanding good skills (level III: 7%): "As far as graduation from upper secondary general education is concerned, the key dividing line runs between proficiency level II and III (the rate is above 10% for the proficiency levels III to V and about 5% for level II and lower)" (Scharenberg et al. 2014, p. 16).

There is another distinctive level on the PISA scale that is considered to represent a kind of minimum standard and is often referred to as the level of *baseline proficiency*. PISA considers level II a baseline level of proficiency at which students begin to demonstrate the reading skills allowing them to participate effectively and productively in life as they continue their studies, and as they enter into the labor market and become active members of society (OECD 2012). It is the key priority for all countries to ensure that as many students as possible attain at least level II, since students scoring below this level struggle to perform many everyday reading tasks and face a disproportionately higher risk of poor post-secondary and labor-market participation (OECD 2010b). The Statistics Canada report on Canada's PISA results described level II as "a baseline of proficiency at which students begin to demonstrate the required competencies to use reading for learning" (Knighton et al. 2010, p. 25). In the TREE study the percentage of students who have not completed an upper secondary program is much higher among low-achievers (proficiency level I or below; between 19 and 37%) than among those who scored in the middle or higher ranges (proficiency levels II to V; between 4 and 10%). Scharenberg et al. (2014, p. 16) conclude that concerning "the attainment of an upper secondary certificate, the major dividing line seems to run between those with (very) low reading literacy skills and those with medium to (very) high skills. Achieving proficiency level II appears to be a minimum requirement in this respect" (see also Stalder et al. 2008).

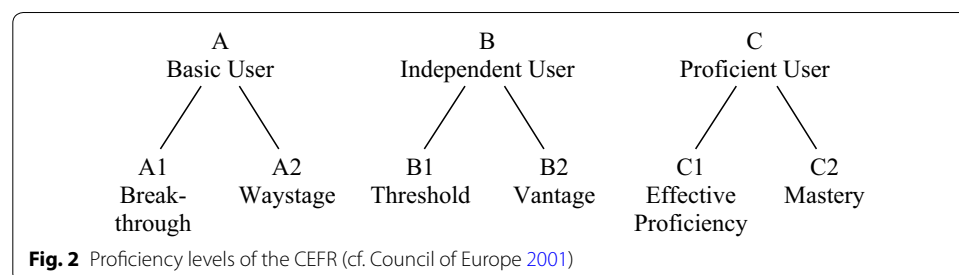### Test of reading comprehension in German NA for non-native speakers

The assessment of English (Expanding Circle, non-native speakers, EFL) reading comprehension with the CEFR-based NA tests is very similar to that of other large-scale studies on basic literacy or reading achievement. In their report on the test development process for NA, Rupp et al. (2008) frequently referred to the PISA study. Essentially, all construct definitions for reading are based on a blend of a cognitive processing, reader purpose, and reading task perspective as compared and contrasted by Enright et al. (2000). German NA measures reading comprehension in English as a foreign language as an active communication skill based on written text that is authentic and considered to be relevant and meaningful by society. As in PISA, the specific purpose of reading comprehension ranges from understanding specific local details to making complex global inferences. Students are required to apply fundamental reading skills and to form mental models of the text that vary in their level of detail, coherence, and complexity (Rupp et al. 2008). Like PISA, German NA uses continuous and non-continuous texts in order to assess a general, literacy-oriented reading ability.

Fleckenstein *et al. Large-scale Assess Educ* (2016) 4:19

Page 8 of 19

In the course of the PISA study some European countries have introduced national assessment programs to monitor educational outcomes. CEFR-based National Educational Standards for English as the first foreign language, for example, were commissioned on behalf of the *Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany* (KMK) and monitored on a regular basis in German NA. German National Educational Standards define EFL as competencies that students are expected to have acquired at a particular grade level or for a certain school-leaving qualification (Rupp et al. 2008). They set normative benchmarks for student achievement in order to establish an accountability system for academic outcomes of lower secondary education ([reference deleted to maintain the integrity of the review process]). These benchmarks are determined in relation to the CEFR levels.

The CEFR was developed to achieve a higher degree of transparency and coherence in language learning and teaching in Europe. It defines productive and receptive levels for different dimensions of language competence and it describes in detail which skills learners have to master in order to communicate successfully in a given language (Council of Europe 2001). The CEFR describes language proficiency in reading, writing, speaking, and listening on a 6-level scale, combined in three superordinate clusters of ascending language proficiency (A, B, and C; see Fig. 2). The six proficiency levels are specified in terms of can-do statements that resulted from a detailed analysis of a number of international scales (North 2000). Since it was published by the Council of Europe in 2001, the CEFR has provided a basis for the planning of examination content and the specification of assessment criteria in many countries (Little 2007). Usually, standard setting procedures are used to map test scores onto the CEFR scale (Cizek and Bunch 2007; Lim et al. 2013).

German National Educational Standards are assessed at the end of lower secondary education (grade 9; [reference deleted to maintain the integrity of the review process]). The scaling procedure used in NA is very similar to the one used in PISA. PVs are drawn from a Rasch model and transformed to a national 500/100 metric. According to their respective score, students are categorized into one of the five proficiency levels (A1-C1; the highest level C2 is not expected to be achieved by EFL students in grade 9).

The objective of this assessment procedure is to observe the extent to which students reach certain politically predefined standards for EFL. A *norm standard* was specified in order to determine the degree of proficiency that most German EFL students should reach. This *norm standard* was located at the upper half of level B1 (B1.2; cut-off score of 550). Performance at or above this cut-off score implies that a student is able to satisfy this particular requirement (compare Table 1). Additionally, there is a *minimum*

Fig. 2 Proficiency levels of the CEFR (cf. Council of Europe 2001)

Fleckenstein *et al. Large-scale Assess Educ* (2016) 4:19

Page 9 of 19

standard specified at level A2.2 with a cut-off score of 450, which should ideally be attained by all students in Germany.

The present study

A global language like English has many definite advantages; however, there are also some potential risks associated with it. Crystal (2003b, p. 16) voiced the concern of many critics of this 'Anglicization' when he asked: "Will those who speak a global language as a mother tongue automatically be in a position of power compared with those who have to learn it as an official or foreign language?" The 'linguistic power' of native speakers can lead to disadvantages for non-native speakers, who compete for positions on the global job market and for admission to higher education. Crystal (2003b), however, is optimistic that this gap can be overcome by effective EFL learning and instruction.

The CEFR is a popular basis for standard-setting procedures such as the one conducted for German NA. But how do we know if and when to call EFL learning effective? What are reasonable expectations for English language proficiency of non-native speakers? The CEFR does not provide normative assumptions about what level of language proficiency should be reached by non-native speakers. As a descriptive framework, it serves the localization of foreign language learners on a scale of can-do descriptors. PISA on the other hand has not performed a systematic standard-setting procedure, which may indeed be criticized. However, PISA and its follow-up studies provide us with an indicator for consequential validity of cut-scores by relating them to certain educational and vocational outcomes. So by using tests from both frameworks, we can draw on the strengths of one to undermine the weaknesses of the other—and vice versa. PISA follow-ups indicate that certain proficiency levels are associated with success later in life, while the CEFR provides a theory-based, widespread framework for localizing students on meaningful levels of language proficiency.

It is a common procedure to localize EFL students on the CEFR in order to compare them with other non-native speakers. We can thereby state what students are able to do and we can declare certain contextualized standards they should be able to achieve. What we cannot know, however, is whether the degree of proficiency we expect of them is what is required in a globalized world where they compete with native speakers from the Outer and Inner Circles. EFL learners are not commonly localized on a global scale of English language competence that is independent of the context in which the language has been learned or acquired. Similarly, national EFL standards are typically not examined in light of global challenges facing EFL students. In order to close this research gap, the present study considers a test for native speakers (PISA) in addition to a test for non-native speakers (German NA). Our specific research questions are the following:

1. Do Expanding Circle students meet the requirements of the Inner Circle?
    a. How competent are German non-native speakers compared to English native speakers?
    b. What percentage of German non-native speakers can be considered proficient according to the PISA English reading literacy scale?
2. How are the proficiency classifications of German NA and PISA as well as corresponding political implications linked to each other?

Fleckenstein *et al. Large-scale Assess Educ* (2016) 4:19

Page 10 of 19

    a. How strong is the relationship between PISA and German NA English reading comprehension scores?

    b. Where are the German NA levels localized on the PISA scale?

    c. Where are the proficiency standards for non-native speakers localized on a native speakers' scale?
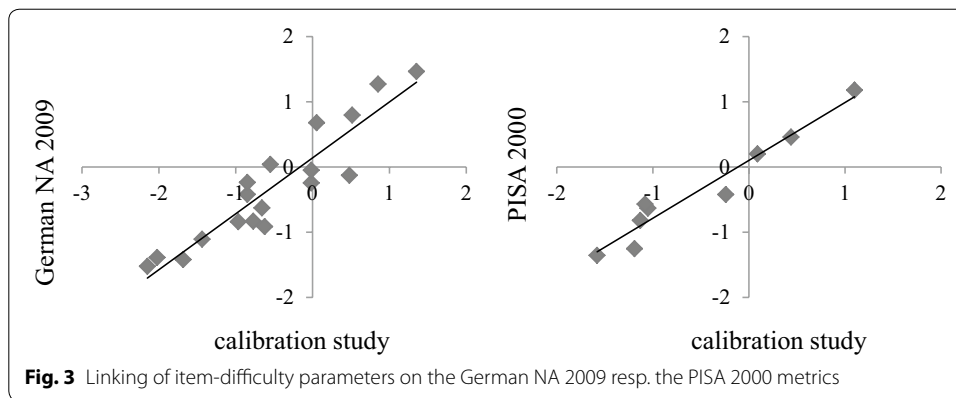
Methods

Sample

The sample comprised 427 students (50.8% female) enrolled in high schools all over Germany. This was a sub-sample of the representative calibration sample for the NA conducted in all German federal states ([reference deleted to maintain the integrity of the review process]). The students' mean age was 15.9 (*SD* = .8) years, ranging from 14.4 to 18.2 years. The students were at the end of compulsory schooling (lower secondary education; ISCED level 2), which in the German school system amounts to either grade 9 (55.7%) or grade 10 (44.3%). Furthermore, the students were subdivided into academic track (*Gymnasium*; 42.4%) and non-academic track (*Hauptschule/Realschule*; 57.6%) students. These tracks in the German secondary school system differ in terms of student achievement levels, years of schooling, and the academic opportunities available after graduation. The academic track prepares students for upper secondary education and university; hence, it is the most selective secondary school in the German school system. The non-academic track usually leads to apprenticeship training and part-time enrollment in higher vocational schools.

In the school year of 2007/2008, 29.8% of all ninth graders and 35.5% of all tenth graders in Germany were enrolled in a Gymnasium (German Federal Statistical Office 2010). These percentages differ from those in our sample, thus, we weighted our data according to the actual distribution of students in the population to enable an international comparison of proficiency levels.

Measures

Data were collected as a part of a calibration study for German National Assessment (NA) 2009. German NA is a national program for the assessment of educational outcomes in different domains at the end of lower secondary education. It is administered regularly every 6 years, starting in 2009 and substituting national PISA supplement studies (PISA-E) that had previously been used in Germany to gather similar information on student achievement. The calibration study was conducted in 2008 to select testlets for the official German NA in 2009. Data were gathered in April and May, 2008. A total of 136 items of the German NA reading comprehension test was administered to the sample in the present study (i.e., a subsample of the calibration study sample). Each student was presented with a subset of items (two blocks of approximately 36 to 42 items) in a balanced incomplete block design. In addition, two released PISA reading literacy testlets (*Runners* and *Lake Chad*; nine items) were administered to all students in the present sample as an external validation criterion.

As the results were to be presented on the official scales of German NA and PISA, the item-difficulty parameters were fixed onto those of PISA 2000 (all nine items) and the subsequent NA 2009 (18 items), respectively (see Fig. 3). They showed a rather strong

Fleckenstein *et al. Large-scale Assess Educ* (2016) 4:19

Page 11 of 19

Fig. 3 Linking of item-difficulty parameters on the German NA 2009 resp. the PISA 2000 metrics

relationship (NA: $R^2 = .86$; PISA: $R^2 = .94$) and were consequently used as anchor items for the scaling procedure. Student abilities were estimated by drawing plausible values (PVs) for all cases in a two-dimensional, Rasch model in ConQuest, Version 3.0 (Wu et al. 2007). Background information (*track*, *grade*) were included as conditioning variables. The PV-reliability was high for both scales (NA: .92/PISA: .84). The mean item difficulty was .61 for German NA and .57 for PISA, respectively. The correlation of the latent variables in the two-dimensional model was $r = .74$ ($p < .001$; 95% CI .69–.78).

For each case, the five PVs were transformed to the 500/100-metrics of PISA 2000 and German NA 2009, respectively, before being recoded into categorical variables according to the corresponding proficiency levels (see Table 1). The resulting proficiency level classifications were subsequently linked to each other in a contingency table. On the basis of this table one can infer what percentage of students localized on a certain NA level is proficient on a certain PISA level. All analyses were performed for the five PVs individually and were then combined in accordance with Rubin (1987).

Results

First of all, results allow for a localization of EFL students on the international PISA scale. Figure 4 shows a comparison of our sample with students from two exemplary Anglophone countries, the UK and the US, in terms of the distribution of proficiency levels. It shows that while there was an approximated normal distribution of student abilities for the first language English students from Anglophone countries, it was positively skewed for our sample of non-native speakers. There were more than twice as many students on PISA level I and, by contrast, only half as many on level IV. The highest proficiency level V was reached by 8.0% resp. 9.9% of native speakers, but only .9% of non-native speakers. The percentages at or above baseline (level II) and norm (level III) proficiency give us an idea of how many students are proficient enough to successfully compete in an English-speaking globalized society: Over 80% of native speakers scored at or above baseline proficiency, in our sample of non-native speakers the percentage was 56.8%. Of the non-native speakers investigated 31.3% reached at least norm proficiency, while the corresponding percentage amongst US-American students was 58.1%.

These results were further differentiated according to the different subgroups of our sample: While the overall mean score was $M = 419$ ($SD = 111$), the mean score for academic track students was $M = 502$ ($SD = 81$). The former differed significantly from the
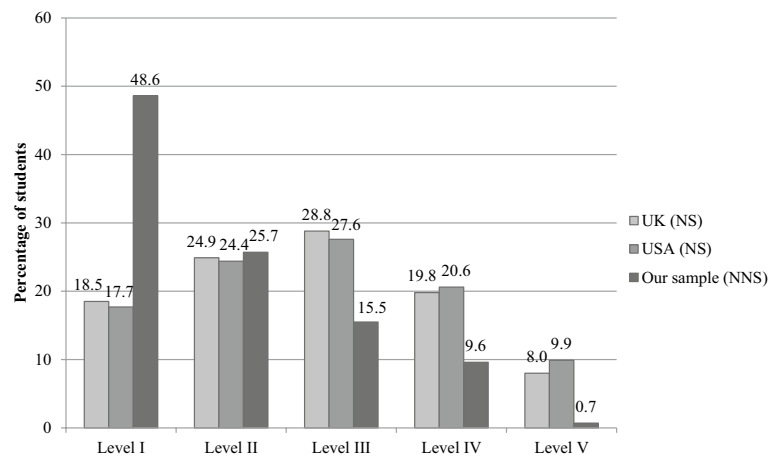
Fleckenstein *et al. Large-scale Assess Educ* (2016) 4:19

Page 12 of 19

Fig. 4 International comparison of the distribution of proficiency levels on the PISA 2009 scale

OECD average in PISA 2009 ($M = 493$; $SD = 93$; $t(426) = -13.89$, $p < .001$, $d = .72$), while the latter did not [$t(180) = 1.51$, $p = .132$, $d = .10$]. The non-academic track students, in turn, achieved a mean score of $M = 357$ ($SD = 87$; $t(245) = -24.59$, $p < .001$, $d = 1.51$). This tendency also became evident in the percentage of students from the different subgroups that reach certain proficiency levels on the PISA scale (see Fig. 5): While two-thirds of academic track students reached norm proficiency, only one-third of non-academic track students scored at or above baseline proficiency.

Secondly, we present a cross-classification of proficiency levels in PISA and German NA which showed a rather strong relationship of $\rho = .65$ (95% CI .59–.70). Table 2 presents the linkage of the proficiency level classifications in order to show where the German NA levels are localized on the PISA scale. Conditional percentages are used to
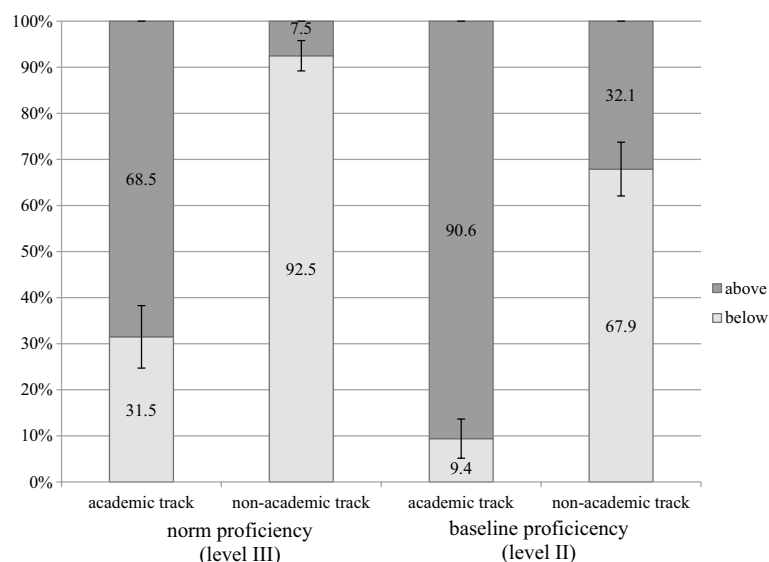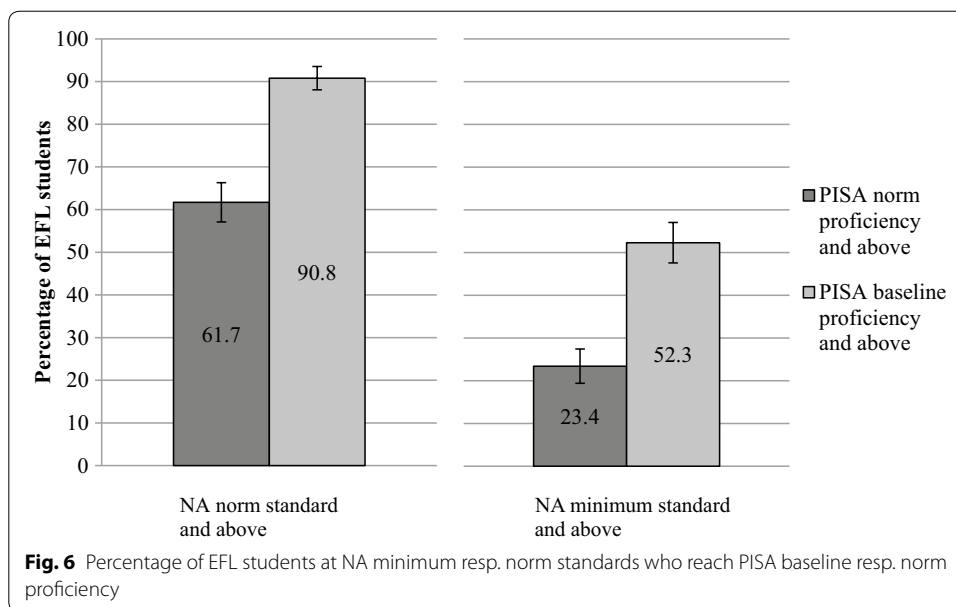

Fig. 5 Percentage of academic and non-academic track students above and below norm and baseline proficiency on the PISA scale

Fleckenstein *et al. Large-scale Assess Educ* (2016) 4:19

Page 13 of 19

Table 2 Contingency table for German NA and PISA (total and percentage in PISA with 95% confidence interval)

	PISA					
	I	II	III	IV	V	Overall
NA						
A1						
Total	32	0	0	0	0	32
% in PISA	*100.0*	0.0	0.0	0.0	0.0	100.0
(95% CI)	*(89.1–100.0)*	(0.0–10.9)	(0.0–10.9)	(0.0–10.9)	(0.0–10.9)	
A2						
Total	105	42	6	3	0	156
% in PISA	67.3	*26.9*	3.9	1.9	0.0	100.0
(95% CI)	(59.3–74.6)	*(20.1–34.6)*	(1.4–8.2)	(0.4–5.5)	(0.0–2.3)	
B1						
Total	44	59	58	17	1	179
% in PISA	24.6	33.0	*32.4*	9.5	0.6	100.0
(95% CI)	(18.5–31.6)	(26.1–40.4)	(25.6–39.8)	(5.6–14.8)	(0.0–3.1)	
B2						
Total	3	8	14	30	3	58
% in PISA	5.2	13.8	24.1	*51.7*	5.2	100.0
(95% CI)	(1.1–14.4)	(6.1–25.4)	(13.9–37.2)	(38.2–65.0)	(1.1–14.4)	
C1						
Total	0	0	1	1	0	2
% in PISA	0.0	0.0	50.0	50.0	*0.0*	100.0
(95% CI)	(0.0–84.2)	(0.0–84.2)	(1.3–98.7)	(1.3–98.7)	*(0.0–84.2)*	
Overall						
Total	184	109	79	51	4	427
% in PISA	43.1	25.5	18.5	11.9	0.9	100.0
(95% CI)	(38.3–47.9)	(21.5–29.9)	(14.9–22.5)	(9.0–15.4)	(0.3–2.4)	

indicate the amount of equal and divergent classifications on each level. We found a systematic shift in the classifications of proficiency levels of the two scales: Most students on a certain German NA level scored at the corresponding or at a lower level on the PISA scale. For example, a student localized on level A2 (second level of the CEFR) is likely to score on PISA level I or II, and a student scoring on level B1 (third level of the CEFR) will probably reach level I, II or III on the PISA scale.

In terms of proficiency standards for non-native speakers (German NA minimum and norm standard) compared to those for native speakers (PISA baseline and norm proficiency), we found a shift that resembles the one found for proficiency levels as shown in Table 2. Figure 6 shows the correspondence of the German NA minimum and norm standard to the PISA baseline and norm proficiency: Of those non-native speakers who attained the German NA norm standard (B1.2 on the CEFR), 61.7 and 90.8% reached norm proficiency and baseline proficiency on the PISA scale, respectively. Of those who reached German NA minimum standard (A2.2 on the CEFR), 23.4% achieved norm proficiency and 52.3% achieved baseline proficiency.

Fleckenstein *et al. Large-scale Assess Educ* (2016) 4:19

Page 14 of 19

Fig. 6 Percentage of EFL students at NA minimum resp. norm standards who reach PISA baseline resp. norm proficiency

Discussion

Summary

The linking of two proficiency classifications, one for non-native speakers and one for native speakers, is a novel approach in language assessment research. We compared two different frameworks which both set their own proficiency standards for different linguistic contexts: non-native speakers in the Expanding Circle and native speakers in the Inner Circle of Kachru's model for English as a world language. On the one hand we used the construct of reading comprehension in German NA. On the other hand we used the construct of reading literacy in PISA to indicate what students are expected to achieve in a world in which English can be considered the global language of communication. The central ideas behind this endeavor were (1) to localize non-native speakers on a scale for native speakers, and (2) to attempt a cross-classification of proficiency levels from two different frameworks for language proficiency.

As expected, compared to native speakers our EFL learners reached lower proficiency levels on the PISA scale. However, there was a discrepancy between the different tracks: While a majority of academic track students reached norm proficiency in PISA, only a minority of non-academic track students even reached baseline proficiency. We could show a systematic shift in proficiency classifications which indicates that a student needs to be more proficient in order to reach the correspondent level on the PISA scale. These results substantially contribute to the validation as well as localization of proficiency standards for non-native speakers on a global scale.

Conclusions

We found a strong latent correlation ($r = .74$) for our two English reading tests which indicates substantial overlap of the constructs that were assessed. However, one might have expected the coefficient to be even higher in order to consider the constructs equivalent. Achievement tests are generally found to correlate strongly: For example,

Fleckenstein *et al. Large-scale Assess Educ* (2016) 4:19

Page 15 of 19

the latent correlation between science and reading literacy in the German PISA 2000 sample was $r = .87$ and the correlation between math and reading was $r = .84$, and science and math were correlated at $r = .83$ (Prenzel et al. 2001). These coefficients indicate relationships between all three domains that are even stronger than the one between our two English reading tests. Thus, the correlation coefficient found in this study may appear rather low, considering the claim of construct equivalence between the two tests. However, in order to appropriately interpret and compare these correlation coefficients, one needs to account for the following: First, reading and comprehending texts plays an important role in all three PISA domains and, in turn, the reading test draws on skills from the other two areas by using graphs and tables. This becomes clear when looking at partial correlations: Once the third domain is controlled for, the correlations drop substantially to $r = .36$ (math/science), $r = .58$ (reading/science), and $r = .49$ (reading/math; Prenzel et al. 2001). Secondly, the variance in the PISA sample was very large as they assessed 15-year-old students from every school type including special-needs students. The inter-correlations within a certain type of school are much lower, ranging from $r = .56$ to $.70$ at the Gymnasium (Prenzel et al. 2001). Since the variance in the sample of the present study is more limited, one would expect the correlations to be weaker. Furthermore, only a small number of items were used to assess PISA literacy in the present study. This limits the breadth of the assessed construct and may lead to an underestimation of the correlation.

Keeping the preceding information in mind, the latent correlation found for our two English reading literacy tests indicates a rather strong relationship and substantial overlap of the two constructs. As the two tests were developed for different contexts and with reference to different frameworks they would not have been expected to correlate perfectly.

To what extent the strength of correlations between two tests from different frameworks depends on differences in construct definitions and test specifications (which, as we were able to show, are rather similar for our two tests) or whether it is the result of actual processing differences demanded by first language versus second or foreign language reading tasks would require a detailed content analysis of the two tests. This could be an interesting topic for future research on the differences and similarities of reading constructs in the first, second, or foreign language, and their operationalization for assessment purposes. Khalifa and Weir (2009) proposed a comprehensive model of the reading process that addresses the role of readers' cognitive operations to enable an empirical investigation of cognitive processing complexity in reading. The model could be very useful in such an endeavour, especially as it explicitly accounts for the operationalization of the reading construct. For a further specification of test content the Dutch Grid—a result of the Dutch CEFR Construct Project (Alderson et al. 2006)—would be a convenient starting point.

Our findings have positive implications for the effectiveness of EFL instruction in German schools—at least in the academic track. Students who learn English in academic track schools achieve levels of reading literacy that are similar to those of average performing students from countries where English is the majority language. This is a success for the academic track in the German secondary educational system. We can say that a German academic track student is probably able to successfully participate in social,

Fleckenstein *et al. Large-scale Assess Educ* (2016) 4:19

Page 16 of 19

economic, and academic English reading literacy environments. However, we found a large discrepancy between the two tracks of schooling: Considering our results it seems rather unlikely for the majority of non-academic track students to be able to compete on the global job market in terms of reading literacy in English.

The results also have implications for national educational policy as we were able to show that non-native speakers' minimum standard and norm standard in NA for EFL actually seem to reflect the skills necessary for success in a globalized English-speaking world. Students who reach the EFL norm standard are very likely to reach at least PISA baseline proficiency. Reaching baseline proficiency means having attained the skills necessary for the effective and productive participation in an English-speaking society. The majority of students who achieve the EFL norm standard even reach PISA norm proficiency, which is a key measure for post-secondary educational academic success in an English-speaking context. If one considers the progressing globalization in terms of education and labor in conjunction with the prominent role that the English language plays in that process, these results are definitely relevant. We can say with some certainty that a student who attains German EFL standards will, later in adulthood, likely be able to participate actively in today's globalized world.

With respect to Kachru's model of the three concentric circles, we are now one step closer to defining the degree of proficiency that non-native speakers from the Expanding Circle have to reach in order to be able to compete and succeed in a world in which English has become the lingua franca. The fact that EFL proficiency classifications seem to be predictive for this goal has positive implications for the validity and relevance of German National Educational Standards.

Limitations and suggestions for further research

Several methodological shortcomings might threaten the validity of our results, including the small number of items on the PISA scale and the lack of external validation criteria. In order to examine the two-dimensionality of the latent constructs English reading in German National Educational Standards and PISA, there should be a similar number of items on both factors. External variables such as the achievement on the German PISA reading literacy test might help interpret the correlation between the two English language constructs. Moreover, the comparison of performance in our sample with PISA results is limited in terms of comparability because of different sampling rationales and different background models: PISA included students with special educational needs, whereas the German NA calibration study did not. Thus, the underperformance of German non-native speakers compared to native speakers may be even more severe than our results indicate. In addition, the background model that was used to draw plausible values (PVs) in PISA could not be fully replicated in the present study. Due to a lack of comprehensive background information on the students, only the two variables with the highest percentage of variance explained were included (school track and grade; $R^2 = .53$).

Another limitation concerns the uncertainty of proficiency level classifications in general: The reduction of continuous test scores to ordinal levels of proficiency necessarily leads to a certain amount of misclassification. It is important to bear in mind that due to measurement sampling error the classification of students into proficiency levels is

Fleckenstein *et al. Large-scale Assess Educ* (2016) 4:19

Page 17 of 19

inevitably deficient (Betebenner et al. 2008). Factors such as test length and the number of proficiency levels may increase the rate of misclassifications of student abilities (Ercikan 2006; Ercikan and Julian 2002). For this reason, Ercikan and Julian (2002) set broad guidelines for desired classification accuracy, depending on the number of levels and the test reliability. For a categorization into five proficiency levels, the authors suggest the following percentages of accurate classifications depending on the assumed reliability of the test (in parentheses): .70 (.90), .60 (.75), .50 (.70). For the official German NA EFL tests, Tiffin-Richards (2011) reported an expected classification accuracy of .76 (.90) on the five CEFR proficiency levels, based on maximum likelihood estimates of examinee ability and their standard errors. This means that in our study the categorization of test scores into five proficiency levels may inherently lead to a misclassification of about 24%, without taking divergent classifications (TOEFL/German NA) into account. This general deficiency of proficiency level classifications, however, does not explain the systematic shift we found in our data.

Additionally, we analyzed data for reading only which disregards the multi-dimensionality of language competence (Jang and Roussos 2007; Leucht et al. 2010). This focus on reading might be one explanation for the high scores especially of academic track students. Reading of different kinds of texts is emphasized in EFL instruction at German schools (Rupp et al. 2008). Thus, it would certainly be interesting to examine other aspects of language competence, such as listening, writing or speaking. Especially speaking skills might suffer because foreign language instruction in an academic setting often does not present students with sufficient opportunities to productively engage with the language (Gilmore 2007). In his account of language proficiency in native speakers and non-native speakers, Hulstijn (2011) claims that second or foreign language learners can reach the same level of higher language cognition (HLC) as native speakers. However, this does not necessarily hold true for basic language cognition (BLC) which includes only verbal skills, namely speaking and listening. A comparison of the verbal skills of native speakers and non-native speakers may lead to very different results than a comparison of reading and confidence in the global participation of EFL students may need to be tempered until there is evidence that students' other language skills also meet an international norm standard. This would certainly be an interesting and relevant topic for future research on the differential outcomes of first or second language acquisition and instructed foreign language learning.

Author details
[1] Leibniz Institute for Science and Mathematics Education, Olshausenstr 62, 24118 Kiel, Germany. [2] Department of Education Studies, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany.

Fleckenstein *et al. Large-scale Assess Educ* (2016) 4:19

Page 18 of 19

References

Alderson, J. C., Figueras, N., Kuijper, H., Nold, G., Takala, S., & Tardieu, C. (2006). Analysing tests of reading and listening in relation to the common European Framework of Reference: The Experience of the Dutch CEFR Construct Project. *Language Assessment Quarterly, 3*, 3–30.

Betebenner, D. W., Shang, Y., Xiang, Y., Zhao, Y., & Yue, X. (2008). The impact of performance level misclassification on the accuracy and precision of percent at performance level measures. *Journal of Educational Measurement, 45*, 119–137.

Birch, B. (2014). *English L2 reading: Getting to the bottom* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Bussière, P., & Knighton, T. (2006). *Educational outcomes at age 19 associated with reading ability at age 15*. Ottawa: Statistics Canada.

Bussière, P., Knighton, T., & Pennock, D. (2001). *Measuring up: The performance of Canada's youth in reading, mathematics and science*. Ottawa: Statistics Canada.

Cha, Y.-K., & Ham, S.-H. (2008). The impact of English on the schools curriculum. In B. Spolsky & F. M. Hult (Eds.), *Handbook of educational linguistics* (pp. 313–327). Malden, MA: Blackwell.

Chall, J. S. (1996). *Stages of reading development* (2nd ed.). Fort Worth, TX: Harcourt-Brace.

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Crystal, D. (2003a). *Cambridge encyclopedia of the English language*. Cambridge: Cambridge University Press.

Crystal, D. (2003b). *English as a global language* (2nd ed.). Cambridge: University Press.

Crystal, D. (2006). English worldwide. In R. Hogg & D. Denison (Eds.), *A history of the English language* (pp. 420–439). Cambridge: Cambridge University Press.

Cummins, J. (1979). Linguistic interdependence and educational development of bilingual children. *Review of Educational Research, 49*, 222–251.

Cummins, J. (1981). *Bilingualism and minority language children*. Ontario: Ontario Institute for Studies in Education.

Cummins, J. (2000). *Language, power and pedgogy: Bilingual children in the crossfire*. Clevedon: Multilingual Matters.

De Zeeuw, M., Schreuder, R., & Verhoeven, L. (2013). Processing of regular and irregular past-tense verb forms in first and second language reading acquisition. *Language Learning, 63*, 740–765.

Dixon, Q., Zhao, J., Shin, J., Wu, S., Su, J., Burgess-Brigham, R., et al. (2012). What we know about second language acquisition. A synthesis from four perspectives. *Review of Educational Research, 82*, 5–60.

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners. *Reading Research Quarterly, 38*, 78–103.

Edele, A., & Stanat, P. (2016). The Role of First-Language Listening Comprehension in Second-Language Reading Comprehension. *Journal of Educational Psychology, 108*, 163–180.

Enright, M. K., Grabe, W., Koda, K., Mosenthal, P., Mulcahy-Ernt, P., & Schedl, M. (2000). TOEFL 2000 reading framework: a working paper. Princeton, NJ: ETS.

Ercikan, K. (2006). Examining guidelines for developing accurate proficiency level scores. *Canadian Journal of Education, 29*, 823–838.

Ercikan, K., & Julian, M. (2002). Classification accuracy of assigning student performance to proficiency levels: Guidelines for assessment design. *Applied Measurement in Education, 15*, 269–294.

Eurobarometer (2012). Europeans and their languages: Special Eurobarometer 386. http://ec.europa.eu/public_opinion/index_en.htm.

Ferguson, G. (2007). The global spread of English, scientific communication and ESP: Questions of equity, access and domain loss. *Ibérica, 13*, 7–38.

Foskett, N. (2010). Global markets, national challenges, local strategies: The strategic challenge of internationalization. In F. Maringe & N. Foskett (Eds.), *Globalization and internationalization in higher education*. London, UK: Continuum.

German Federal Statistical Office (2010). Bildung und Kultur – Allgemeinbildende Schulen. Schuljahr 2007/2008, Fachserie 11, Reihe 1. Wiesbaden, Germany: Statistisches Bundesamt. https://www.destatis.de/GPStatistik/receive/DEHeft_heft_00005576.

Gil, J. A. (2011). A comparison of the global status of English and Chinese: Towards a new global language? *English Today, 27*, 52–59.

Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teacher, 40*, 97–118.

Grabe, W., & Stoller, F. L. (2013). Teaching reading for academic purposes. In M. Celce-Murcia, D. M. Brinton, & M. A. Snow (Eds.), *Teaching English as a second or foreign language* (4th ed.). Boston: Heinle Cengage.

Graddol, D. (2006). *English next*. London: The British Council.

Hu, G. W. (2007). The juggernaut of Chinese-English bilingual education. In A. Feng (Ed.), *Bilingual education in China: Practices, policies and concepts* (pp. 94–126). Clevedon, UK: Multilingual Matters.

Hulstijn, J. H. (2011). Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly, 8*, 229–249.

Jang, E. E., & Roussos, L. (2007). An investigation into the dimensionality of TOEFL using conditional covariance-based nonparametric approach. *Journal of Educational Measurement, 44*, 1–21.

Jenkins, J. (2014). *English as a lingua franca in the international university: The politics of academic English language policy*. Abingdon, UK: Routledge.

Jongejan, W., Verhoeven, L., & Siegel, L. (2007). Predictors of reading and spelling abilities in first-and second-language learners. *Journal of Educational Psychology, 99*, 835–851.

Kachru, B. B. (1986). *The alchemy of English*. Oxford, NY: Pergamon.

Kachru, B. B. (1988). The sacred cows of English. *English Today, 16*, 3–8.

Kachru, B. B. (1992). World Englishes: Approaches, issues and resources. *Language Teaching, 25*, 1–14.

Kachru, B. B. (2011). World Englishes and English-using communities. *Annual Review of Applied Linguistics, 17*, 66–87.

Khalifa, H. & Weir, C. (2009). *Examining reading: Research and practice in assessing second language reading*. Studies in Language Testing, vol. 29. Cambridge: Cambridge ESOL & Cambridge University Press.

Fleckenstein *et al. Large-scale Assess Educ* (2016) 4:19

Page 19 of 19

Kintsch, W., & Mangalath, P. (2011). The construction of meaning. *Topics in Cognitive Science, 3*, 346–370.

Knighton, T., Brochu, P., & Gluszynski, T. (2010). *Measuring up: Canadian results of the OECD PISA study: The performance of Canada's youth in reading, mathematics and science: 2009 first results for Canadians aged 15*. Ottawa: Statistics Canada.

Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. Cambridge, NY: Cambridge University Press.

Köller, O., Knigge, M., & Tesch, B. (Eds.). (2010). *Sprachliche Kompetenzen im Ländervergleich. [Language competencies in German National Assessment]*. Münster: Waxmann.

Leucht, M., Retelsdorf, J., Möller, J., & Köller, O. (2010). Zur Dimensionalität rezeptiver Kompetenzen im Fach Englisch [On the Dimensionality of Receptive Skills in English as a Foreign Language]. *Zeitschrift für Pädagogische Psychologie, 24*, 123–138.

Lesaux, N. K., Rupp, A. A., & Siegel, L. S. (2007). Growth in reading skills of children from diverse linguistic backgrounds: Findings from a five-year longitudinal study. *Journal of Educational Psychology, 99*, 821–834.

Lim, G. S., Geranpayeh, A., Khalifa, H., & Buckendahl, C. W. (2013). Standard setting to an international reference framework: Implications for theory and practice. *International Journal of Testing, 13*, 32–49.

Little, D. (2007). The Common European Framework of Reference for languages: Perspectives on the making of supranational language education policy. *The Modern Language Journal, 91*, 645–653.

Maringe, F., & Foskett, N. (2010). Introduction: Globalization and universities. In F. Maringe & N. Foskett (Eds.), *Globalization and internationalization in higher education* (pp. 1–13). London: Continuum.

Martin, C. D., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language, 69*, 574–588.

North, B. (2000). *The development of a common framework scale of language proficiency*. New York: Peter Lang.

OECD. (2003). *Literacy skills for the world of tomorrow: Further results from PISA 2000*. Paris: OECD Publishing.

OECD. (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: OECD Publishing.

OECD. (2009). *PISA 2009 assessment framework: Key competencies in reading, mathematics and science*. Paris: OECD Publishing.

OECD. (2010a). *Pathways to success: How knowledge and skills at age 15 shape future lives in Canada*. Paris: OECD Publishing.

OECD. (2010b). *PISA 2009 results: What students know and can do: Student performance in reading, mathematics and science* (Vol. 1). Paris: OECD Publishing.

OECD. (2012). *Learning beyond fifteen: Ten years after PISA*. Paris: OECD Publishing.

Prenzel, M., Rost, J., Senkbeil, M., Häußler, P. & Klopp, A. (2001). Naturwissenschaftliche Grundbildung: Testkonzeption und Ergebnisse [Basic science competencies: Test conception and results]. In: J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, & W. Schneider et al. (Eds.), PISA 2000. *Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 192–250). Opladen: Leske + Budrich.

Romaine, S. (2006). Global English: From island tongue to world language. In A. van Kemenade & B. Los (Eds.), *The handbook of the history of English* (pp. 589–608). Malden: Blackwell Publishing.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance* (Vol. 6). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rupp, A. A., Vock, M., Harsch, C., & Köller, O. (2008). *Developing standards-based assessment tasks for English as a first foreign language – Context, processes, and outcomes in Germany*. Münster: Waxmann.

Scharenberg, K., Rudin, M., Müller, B., Meyer, T. & Hupka-Brunner, S. (2014). *Education pathways from compulsory school to young adulthood: The first ten years*. Results of the Swiss panel survey TREE, part I. Basel: TREE.

Stalder, B. E., Meyer, T., & Hupka-Brunner, S. (2008). Leistungsschwach - Bildungsarm? PISA-Kompetenzen als Prädiktoren für nachobligatorische Bildungschancen. [Underperforming – undereducated? PISA competencies as predictors of post-obligatory educational options]. *Die Deutsche Schule, 100*, 436–448.

Svartvik, J., & Leech, G. (2006). *English: One tongue, many voices*. Houndmills: Palgrave Macmillan.

Tiffin-Richards, S. P. (2011). *Setting standards for the assessment of English as a foreign language: Establishing validity evidence for criterion-referenced interpretations of test-scores*. Berlin: Freie Universität. **[Dissertation Thesis]**

Urquhart, A. H., & Weir, C. J. (2013). *Reading in a second language: Process, product and practice*. New York: Routledge.

Wu, M. L., Adams, R. J., Wilson, M., & Haldane, S. A. (2007). *ACER ConQuest. Version 2.0. Generalised item response modeling software* [Computer software]. Camberwell: ACER Press.