Aiding culturally responsive assessment in schools in a globalising world

Guri A. Nortvedt[1] · Eline Wiese[1] · Martin Brown[2] · Denise Burns[2] ·
Gerry McNamara[2] · Joe O'Hara[2] · Herbert Altrichter[3] ·
Magdalena Fellner[3] · Barbara Herzog-Punzenberger[4] · Funda Nayir[5] ·
Pervin Oya Taneri[6]

Abstract

Across the world, teachers' classroom assessment tasks and responsibilities are becoming more diverse due to increased migration. In this review, we address how migrant students are affected by assessment, both summative and formative, at the classroom level, with a focus on culturally responsive assessment (CRA). Previous research has shown that CRA practices mainly occur in student-centred classrooms. Furthermore, both student and teacher beliefs about teaching and learning might negatively affect migrant students' opportunities to engage in assessment situations. Teaching and assessment practices should be negotiated and aligned with and included in classroom norms to be culturally responsive. We propose that what is generally considered a valid and reliable assessment practice might need to be adjusted to account for students' cultural ways of knowing and participating and how this is expressed and communicated within the classroom.

Keywords Migrant students · Culturally responsive assessment · Diversity · Classroom assessment · Beliefs · Validity

---

✉ Guri A. Nortvedt
g.a.nortvedt@ils.uio.no

1   University of Oslo, Oslo, Norway

2   Dublin City University, Dublin, Ireland

3   Johannes Kepler University Linz, Linz, Austria

4   University of Innsbruck, Innsbruck, Austria

5   Pamukkale University, Denizli, Turkey

6   Cankiri Karatekin, Çankırı, Turkey

1 Introduction

Migration patterns have changed in recent years, which have increased the diversity of communities across Europe (Vertovec 2007, 2010). Coupled with the notion of 'education for all' (cf. UNESCO 1994), this presents educators with numerous challenges in meeting the educational needs of migrant students, as the schools encounter issues related to settling in (Smyth et al. 2009), racism (Devine et al. 2008; Rousseau and Tate 2003), early school leaving, and academic success (OECD 2015). These issues indicate that educators are struggling to meet the demands of a diverse classroom. Another indication of this struggle is the achievement gap between majority and migrant students identified in previous research (e.g. Lemu 2015; OECD 2015), with migrant students frequently scoring below majority students. Furthermore, greater differences have been observed between first-generation migrant students and majority students than between second-generation migrant students and majority students (OECD 2015).

While achievement gaps have traditionally been attributed to migrant students' deficits or lack of competence, an Organisation for Economic Co-Operation and Development (OECD) analysis of Programme for International Student Assessment (PISA) 2012 mathematics achievement data revealed that students who had migrated from the same geographical areas performed very differently on the PISA assessment, depending on the country in which the students were now living and attending school, even when controlling for socioeconomic status, or SES (OECD 2015). Two likely interpretations of this outcome are that either (1) the school system or the teachers in the different countries are not equally prepared to handle diversity, or (2) systematic differences might exist in the educational services that are offered, such as inequities in the quality of learning opportunities available for students from immigrant-background families compared to those available to other students. This situation could be related to the structure of the school system and its orientation towards inclusiveness, training provided to teachers and their handling of diversity, or other educational offerings and services (see also Stobart 2005). The situation may even be related to curriculum or the presence (or absence) of inclusive integration policies that promote intercultural approaches (Arikan et al. 2017). According to Bilgili et al. (2018), 'immigrant students' education can be viewed as a factor and indicator of integration' (p. 5). As such, achievement gaps also indicate that work remains to be done to ensure that teaching and assessment practices enhance equity and equality in schools for both migrant and majority students.

Various factors contribute to student learning, such as teacher and instructional quality (Hattie 2009). The relationship between assessment and student learning has received increased focus in the past decade, with emphasis on the relationship between feedback and learning (e.g. Black and Wiliam 2012a; Gardner 2012; Hattie and Timperley 2007). While the focus on 'education for all' emerged some time ago, the focus on equity issues and inclusion has intensified, and researchers have now turned their attention to how assessment affects migrant students (e.g. Heritage and Wylie 2018; Kim and Zabelina 2015; OECD 2015). Related to this area of study is the emergence of the concept of culturally responsive assessment or CRA (Qualls 1998; Slee 2010).

Students bring to school not only their previous knowledge and experiences but also their cultural ways of engaging and communicating (Hodge and Cobb 2016;

Moschkovich and Nelson-Barber 2009). Because migrant students are a diverse group, teachers may need to adapt their assessment or teaching practices to meet their students' varying needs. For example, what constitutes valid and desirable knowledge differs both within and across educational contexts (e.g. Solano-Flores and Nelson-Barber 2001; Stobart 2005) and also between migrant students and their teachers, which may adversely affect the students. Classroom-level assessment practices, thus, must be examined to understand how such practices might affect migrant students' opportunities for learning. For instance, they may not understand what competence they are required to demonstrate in assessment situations. In these cases, differences in what is viewed as valid knowledge can pose a threat to the reliability of the assessment instrument applied and, consequently, can lead to questions about the instrument's validity.

The aim of this article is to review research on the assessment of migrant students, with a focus on CRA in compulsory education in classrooms with migrant students. Based on the review outcomes, we discuss how current and past assessment practices affect migrant students and what CRA might entail in diverse classrooms. The research question we seek to answer is as follows: How can classroom-based assessment that acknowledges and respects learners' cultural backgrounds and approaches to learning be developed and supported?

2 Methodology

Our aim with this article is to review research literature that has been published primarily in the English language in international peer-reviewed journals and books to identify challenges and obstacles to the classroom assessment of migrant students. For the purpose of this study, we define migrant students as those born abroad or whose parents were born abroad (e.g. Bilgili et al. 2018; OECD 2015). Migrant students are a heterogeneous group that comprises students from diverse social and cultural backgrounds and includes, for instance, refugees as well as children from families who have migrated for work and educational purposes. Much of the relevant research that has focused on challenges in assessment has addressed minority students or second language learners, thus including students with a migration background. As such, we have included literature on minority students in the current literature search to allow us to tease out the 'wicked problems' of the educational assessment (EA) of migrant students.

This literature review can be characterised as a state-of-the-art review (Grant and Booth 2009) since we focus on both the current state of knowledge and past issues of relevance to identify what might constitute CRA in a European setting; we also point to emergent matters to set priorities for future research and investigations.

We used the Oria[1] search engine to search several databases, using the terms 'migrant' and 'minority students' in combination with 'culturally responsive

[1] Oria is a search engine supported by the Norwegian University Library that searches several external databases simultaneously (https://bibsys-almaprimo.hosted.exlibrisgroup.com/primo-explore/search?institution=UBO&vid=UIO&tab=default_tab&search_scope=default_scope&mode=Basic&displayMode=full&bulkSize=10&highlight=true&dum=true&displayField=all&lang=no_NO&query=). Searches are performed in a large number of library databases, including Springer, Thomson Reuter and ProQuest.

assessment', 'achievement', 'classroom assessment', 'teacher assessment' and 'assessment for learning'. The searches returned a large number of articles: for instance, 'migrant' + 'culturally responsive assessment' returned 1623 articles, and 'migrant' + 'achievement' returned 21,567 articles. Adding the term 'compulsory education', 'secondary education' or 'primary education' significantly reduced the number of hits. For example, adding 'compulsory education' to the first two searches mentioned reduced the number of hits to 148 and 2527, respectively. Overlap was found in the two searches, and some articles addressing other scientific disciplines were found. Due to the complexity of the content matter, a systematic search could not be undertaken, so less emphasis was put on the initial searches. Based on qualitative analysis and evaluation of the content, we scanned the search results to identify articles and book chapters relevant to assessing migrant students and the associated challenges and obstacles. In total, the analysis included 103 contributions, most of which were peer-reviewed articles, along with some peer-reviewed book chapters and three reports. Several contributions (30) addressed educational assessment on a more general basis. Many of the issues discussed are not novel—some were identified more than 20 years ago—but more recent research allowed for a more nuanced and in-depth view of the challenges in question. A few contributions targeted subtopics, such as teachers' beliefs about students who withdraw or who are passive during assessment situations. We found literature on these subtopics by following discussions on themes that had emerged from articles we had identified in the initial search.

The tradition of multicultural education research emerged in Australia, Canada and the USA as a response to educational systems that excluded black, Hispanic and native (American and Aborigine) students and their cultures, knowledge and perspectives (e.g. Brentnall and Hodge 1984; Cahill 1986; Lingard 1983; Mazurek and Kach 1983; Murphy 1986; Reagan 1984). As such, much of the prior research was conducted in English-speaking communities. In addition, many classroom studies included in the literature review addressed bilingual classrooms or classrooms that included native students. In these cases, the bilingual or native students spoke the same language and shared cultural references and backgrounds. The findings of these studies may have limitations regarding their relevance for European classrooms today, as current European classrooms are more diverse and multilingual (Herzog-Punzenberger et al. 2017), where few or no students share the same language or the same cultural references and backgrounds. Consequently, some of the advice provided in the reviewed research, such as learning the home language of students, is not feasible.

For the current European situation, a discussion is needed on what CRA might entail in educational contexts in which different migration patterns exist and where more diversity is present at the classroom level. Because assessment for learning (AfL) has a strong hold in many European countries, whether and how AfL might be integrated into CRA practices is another important question. In this article, we, therefore, draw on research related to challenges in assessment practices that aim to overcome cultural barriers in classroom assessment. We also review research that points to promising practices to enhance equity in assessment situations in order to identify future research and teaching priorities.

3 Culture, diversity and migration

With increased migration, classrooms have become more culturally complex. The word 'culture' itself is a multifaceted term that can be used and defined in several ways. As described by the OECD (2016, p. 7), '"culture" is difficult to define because cultural groups are always internally heterogeneous and contain individuals who adhere to a range of diverse beliefs and practices'. In this article, we view culture as a product of social interaction, in terms of how culture affects the individual's cognitive construction of knowledge and ways of knowing and of how culture can serve as a basis for identity formation and social participation constructed through participation in different social groups and discourses (Frierson et al. 2010; Stephens 2007). In line with the OECD (2016) definition, culture is viewed as dynamic, fluid and changing, which implies that people's identities, which are constructed in their interactions or affiliations with different cultural constructs or understandings, are also changeable and changing.

Identity in this case, therefore, is not stagnant but is discursive and constructed through the social and cultural contexts that individuals encounter and in which they participate. The impact that different cultures have on identity will 'depend on the extent a social context focuses on a particular identity, and on the individual's needs, motivations, interests and expectations within that situation' (OECD 2016, p. 7). As shown by Bernstein (2000) and Bourdieu (1984, 1990), among others, expressions of culture (i.e. cultural traits) are not equally valued in society nor in school. Through unequal classification and valuation of the individual's culture, knowledge, beliefs and values, schools create social hierarchies. This situation then affects how students' knowledge is valued in the school and, therefore, also affects how students are assessed, as well as the feedback and even the grades they receive.

Vertovec (2007, 2010) coined the term 'super-diversity' to describe diversity in contemporary society. The term points to a complexity within modern societies that has occurred through changes in migratory patterns over the last 30 to 40 years, among other factors. According to Vertovec (2007, 2010), the differences and diversity experienced now surpass past experiences and have led to diversity not only in religion, country of origin and language but also in the social rights and status achieved by different immigrant groups who have arrived at varying times and with different social statuses. The result is that differences have increased among those who were previously viewed as one group (for example, Pakistanis in the UK), so these new groups or formations now must be viewed separately.

As described by Arnaut, Blommaert, Rampton and Spotti (2016, p. 2), 'there has been radical diversification not only in the socioeconomic, cultural, religious, and linguistic profiles of the migrants, but also in their civil status, their educational or training background, and their migration trajectories, networks, and diasporic links'. This increase in differences has led to a diversification of society in which former classifications and social stratifications now seem insufficient to describe the complexity that society currently experiences (Meissner and Vertovec 2015; Vertovec 2007, 2010). At the classroom level, this increased complexity necessitates teacher knowledge and awareness of student backgrounds. Teachers must also be sensitive to how these factors might interact and how they might affect participation and learning and, thus, assessment situations and practices.

4 Assessment in diverse classrooms

Migrant students form a heterogeneous group, with diverse backgrounds, legal statuses and rights (Bilgili et al. 2018), all of which can affect assessment in schools. Language, ethnicity, country of origin, religion, gender and social or economic status or class differences often arise in current political and educational debates (e.g. Arikan et al. 2017: Klinger et al. 2018; Stobart 2005). Understanding differences in such factors as students' values, interests, competencies and experiences is relevant for classroom learning and assessment. According to the US National Research Council (NRC 2001, p. 1), 'educational assessment seeks to determine how well students are learning and is an integral part of the quest for improved education. It provides feedback to students, educators, parents, policymakers, and the public about the effectiveness of educational services'. As such, EA can serve multiple purposes and functions (Eder et al. 2009).

First, in its educational function, the aim of EA is to facilitate learning (Cook 1951) by providing feedback to students, teachers and parents, by motivating further effort and by signalling what is important. Second, in its societal function, EA provides the basis for certification (and, as a consequence, for the selection of students and their allocation in the labour market and society) and for informing employers and further education providers about the competencies students have acquired. It also informs education policy about the performance of the education system. In these situations, while EA should, in theory, allow migrant students to demonstrate their competences to the same extent as majority students. The response patterns identified in international large-scale assessments have indicated that this is not always the case (Arikan et al. 2017; OECD 2015). Third, EA may also have personal consequences, as the practice may strengthen or undermine students' self-concepts, attitudes towards learning and continued educational aspirations (Aronson and Laughter 2016; Moschkovich 2007). Previous research has identified challenges related to all three functions of EA with respect to migrant students (e.g. Civil and Hunter 2015; Hattie and Timperley 2007; OECD 2015; Stobart 2005).

Educational assessment may be understood as a composite process: the act of eliciting information about students' learning and/or competencies from their work on appropriate tasks is complemented by ascribing an informative or evaluative judgement through a comparison of student learning outcomes with performance criteria. Either phase of EA may be done internally (e.g. teacher assessment) or externally (e.g. national tests and exams). Students may also contribute to EA, through either self- or peer-assessment (Hayward 2012; Heritage and Wylie 2018). Educational assessment should be based on a set of scientific principles and philosophical assumptions that are grounded in an understanding or theory about how people learn, what they know and how knowledge and understanding develop (Brookhardt 2009; Gardner 2012; Shepard 2006).

Any assessment should be designed based on knowledge or theoretical assumptions about which kinds of tasks are most suitable for eliciting demonstrations of important knowledge and skills from students. Assessment outcomes, or evidence, should be carefully interpreted to draw meaningful inferences about what students know and can do (Gardner 2012; NRC 2001). Ideally, students should learn something important in every assessment situation, and researchers have long argued that we need to move

towards high-quality tasks that will elicit important information (e.g. Burkhardt and Schoenfeld 2018). However, the process of designing, conducting and interpreting assessments is complex, in particular, when assessing complex skills, such as critical thinking and problem-solving (Gipps and Murphy 1994; Klenowski 2009; Siemon et al. 2004; Stobart 2008). Designing assessment situations and tools that are culturally fair to all students adds to the complexity. Furthermore, alignment between curricula, teaching practices and assessment tasks is also necessary for the assessment situation to be valid and transparent (Ayalon and Livneh 2013; Heritage and Wylie 2018; Stobart 2005) in all classrooms, especially in diverse classrooms. For instance, student learning that results from being exposed to classroom practices that are oriented towards problem-solving and the use of open-ended problems should not be assessed using only routine tasks.

4.1 Culturally responsive assessment

The literature on culturally responsive issues originated in the tradition of multiculturalism and addresses learning and pedagogy (Brown-Jeffy and Cooper 2011) in bilingual or bicultural classrooms. This research is predominantly concerned with both matters of effectiveness (e.g. equality in learning outcomes) and ethics (e.g. minority rights and how the majority interacts with cultural minority groups). Following increased concerns about equity in education and inequalities in the educational achievement of minority and migrant students compared to that of majority students (e.g. Hood 1998a; 1998b), focus has been placed on culturally responsive teaching and learning that encompasses the diversity of contemporary classrooms (e.g. Bradshaw et al. 2008; Moschkovich and Nelson-Barber 2009; Raines et al. 2012; Solano-Flores and Nelson-Barber 2001).

Culturally responsive assessment should consider students' cultural ways of communicating and acting within and outside the classroom (Hood 1998b; Kirova and Hennig 2013; Klenowski 2009). Furthermore, CRA is student- and culture-centred (Ford and Kea 2009). As such, CRA requires that educators respect students' perceptions of culture and their way of identifying with peers and teachers alike. Having respect for different aspects of culture could mean, for example, respecting students' preferences to learn in a group; in such cases, the students may seek collective rather than individual achievement (Moschkovich and Nelson-Barber 2009). Respect for students' cultural differences also entails being aware of whether students consider knowledge to be something that belongs to the collective or to the individual. Furthermore, teachers must also be cognisant of classroom activities that conflict with students' culturally based ways of participating (Cumming and van der Kleij 2016), such as a student's preference not to be asked for a personal response or to critique another student's reasoning for fear of being offensive (Civil and Hunter 2015; Hunter et al. 2016). Teachers can facilitate migrant students' participation in assessment situations in different ways, such as by implementing diverse assessment and response formats and by adhering to student authority as well as cultural ways of communicating. Consequently, CRA refers to assessment designs, processes and adaption to individual students and classrooms as well as to assessment outcomes that are sensitive to cultural variations in ways of participating, thinking and learning.

In order to achieve a fair or culturally responsive assessment, increased cultural competence in schools must be achieved (Gay 2018). For students from migrant backgrounds, the outcomes of the assessment are significant for their future learning and development, just as for other students. However, the equity issues discussed previously have implications for teachers' convictions and practices. Teachers need to develop cultural competence and awareness through reflective practice (AEA 2011), which requires them to be aware of the social, political, economic and structural inequalities between the students in their classrooms (Cummins 2015). Cummins maintained that culturally responsive education and assessment at the school level challenges 'patterns of discrimination and exclusion' and 'promotes academic achievement and equality of educational opportunity' (2015, p. 457).

4.2 Classroom-level CRA

Classroom assessment (CA) is assessment used by teachers to determine whether students have mastered the body of knowledge, skills or strategies targeted in the teaching (Brookhardt 2009). Classroom assessment is sometimes referred to as teacher assessment, as it is primarily (although not always) designed and conducted by teachers. In addition, CA may include external classroom-level assessment (e.g. national tests), where the teacher has access to and can utilise the assessment outcomes as evidence of student learning or to plan classroom teaching or interventions; CA may include self- and peer-assessment performed by students as well (Hayward 2012). According to Brookhardt (2009), quality CA rests on a clear understanding of learning goals and an alignment between learning goals and assessment formats. Classroom assessment should include feedback and opportunities for using that feedback to enhance learning. Shepard (2006) claimed that both summative and formative assessments are part of CA. A key question is how CA practices might become culturally responsive.

Summative assessment may be understood as assessment that seeks to elicit information that can be used to determine whether a student has learned specific content or has achieved a particular learning objective by a given point and, as such, to verify the student's attainment of important learning goals (Shepard 2006). The term 'summative assessment' is often used when the intention is to grade student work (Brookhardt 2009). Teacher-made tests are often used in these cases. For summative assessment to be culturally responsive, the applied assessment needs to elicit high-quality information from all students, regardless of their background. Previous research has shown, however, that the purposes and strategies of summative assessment do not always reflect an awareness of students' cultural ways of interacting, as they are typically designed for the majority population (Basterra et al. 2011; Pollitt et al. 2000; Stobart 2005). Again, this situation lessens students' opportunities for demonstrating their competences and brings into question the validity of the assessment outcomes.

Previous research has indicated that teachers may have difficulty designing assessment formats that elicit evidence of complex skills, understanding or creative thinking and may, instead, primarily design tasks that elicit evidence of factual knowledge or trained skills (e.g. Boesen 2006). One might, therefore, argue that this scenario is also especially important to consider when designing tasks for students with cultural and

linguistic backgrounds and prior knowledge that differ from the background of the majority students.

In discussing the term 'formative assessment', Sadler (1989) stated that teachers should provide feedback to students on the correctness of their work and should provide a link to performance standards to facilitate learning; students should also be provided with strategies for reaching these standards. Several years later, the British Assessment Reform Group (Gardner 2012) introduced the term AfL, where the focus is on learning. In AfL, assessors (students or their teachers or peers) use the insights provided by the assessment activity to enhance student learning through clear and appropriate feedback (Black and Wiliam 2012b).

The introduction of AfL not only changed the purpose of assessment but also brought to the forefront student involvement in assessment practices, clarity regarding assessment criteria and expectations and high-quality assessment procedures. Several reviews have indicated that AfL may have a major impact on student learning (Black and Wiliam 1998; Hattie 2009; Hattie and Timperley 2007; Köller 2005; Leahy and Wiliam 2012; Wiliam 2007), provided the assessment satisfies certain defined criteria. First, the assessor must identify and correctly interpret evidence of student knowledge and skills, and the evidence must be relevant for the intended feedback. Second, the feedback must be focused on short-term learning. Finally, the student must understand and be capable of following up on the feedback.

According to Heritage and Wylie (2018), AfL can contribute to equity and learning for minority students, as this approach to assessment allows teachers to both recognise differences between students in their classrooms and respond to their varying needs. As such, AfL may contribute to CRA-oriented practices, although this may require teachers to create new models of assessment compared to more traditional, less student-oriented practices.

The evidence that teachers use to provide feedback can come from a range of assessment formats and procedures: for instance, from a test, by observing student work, or from a portfolio or artistic performance. As such, the divide between summative and formative assessment becomes somewhat blurred, since individual teachers can even use data from national tests, mainly developed to monitor student learning, to provide feedback and design learning activities for individual students. With AfL, the roles of teachers and students change compared to the roles they play in traditional education, where teachers have the main responsibility for the teaching. In AfL classrooms, teachers and students ideally share responsibility for learning, thus showing a shift from a teacher-centred to a student-centred focus on teaching, learning and assessment (Heritage and Wylie 2018). This shift necessitates changes in both teachers' and students' thinking or meta-cognitive awareness regarding assessment and classroom discourse. In particular, the active identification of students' prior and current knowledge and skills is necessary, which is also central in CRA.

Practising AfL is challenging, and across Europe, researchers reported that teachers have not yet learned to master AfL practices (Nortvedt et al. 2016; Schmidinger et al. 2015). Previous research has pointed to possible tensions between having a strong focus on accountability and the use of more classroom-centred assessment traditions like AfL (Stobart 2008, 2012). We find that if the conditions of AFL are fulfilled, if assessment becomes more student-centred and if teachers are cognisant of cultural ways of learning and knowing, then AFL practices might align well with the principles

of CRA. Even so, certain challenges must be overcome in terms of the assessment of migrant students if CA is to become truly culturally responsive.

5 Challenges in assessing migrant students

Several challenges have been identified in the research literature regarding the assessment of migrant students, particularly related to the issues of achievement standards; assessment practices and formats; alignment between teaching, learning and assessment; and various factors connected to the teachers or students themselves. Although these issues may appear to be separate, in reality, they are intertwined. In our analysis of the previous research literature, we attempted to disentangle common sources of the challenges that run across these issues. For instance, education policy regulates assessment standards and influences the alignment of curriculum and assessment, the relation between formative and summative assessment practices and the assessment procedures teachers apply in their classrooms. As such, education policy is a potential source of some of the challenges to assessing migrant students. Similarly, teachers' beliefs affect how they interpret education policy, how they interpret education standards, what they value in the curriculum and how they implement AfL. These factors may all affect the opportunities migrant students have to participate in assessment situations in their classrooms.

Based on these common sources of challenges to the assessment of migrant students, we identified four broad themes in our review: challenges related to (1) policy and standardisation, (2) teachers' and students' beliefs and conceptions of competence, (3) communication and participation and (4) student background. Each factor influences the validity and reliability of classroom-level assessment. For instance, teachers' interpretations of curriculum and education policies influence both what they include in assessment situations and the degree to which they adapt the situation to accommodate student participation. If the curriculum emphasises critical thinking but the teacher only offers students opportunities to demonstrate factual knowledge, then the validity of the assessment will be compromised.

5.1 Challenges related to policy and standardisation

The level of standardisation in education differs from country to country (Klinger et al. 2018). Policies developed at the national level can aim to assist schools in managing diversity. However, according to Klinger et al. (2018), few policies exist that explicitly target the education of migrant students. Although education policies may be purposefully designed to narrow the achievement gap between majority and migrant students, such policies can have unintended consequences and may even increase the gap (Bilgili et al. 2018). The Migrant Integration Policy Index (MIPEX 2015) is used to assess how open and supportive a country is towards its migrant population through the analysis of various policy dimensions. Appreciating cultural diversity and providing access to education are among the success factors. In the 'TIES' study, an eight-country comparative study on the educational careers of second-generation students (conducted after they had completed their education) major differences in levels of academic success was found (Crul et al. 2012). Following the integration-context theory (Crul and Schneider

2010), the structure of the educational system seems to have an important impact on those levels, as do immigration and integration policies. Schnell (2014) identified the age when children usually begin early childhood education, the age when students are tracked into different school types (academic vs vocational focus), the hours they spend in school and the overall selection or inclusion logic of the school system (e.g. repetition rates and overall differentiation) as key characteristics for the academic success of the descendants of immigrants. Bilgili et al. (2018) found that a high level of differentiation ('tracking' or 'streaming') in the school system was associated with a negative effect on migrants' educational achievement and that a more comprehensive or only moderately differentiated system would likely benefit migrant students.

Analysis of PISA data suggests that the performance differences between migrant and majority students have shrunk over time (OECD 2015). It should be noted that Klinger et al. (2018) argued that these differences might be greater than the OECD analysis indicates since large-scale international studies are typically insensitive to change. As such, more targeted policies or changes towards a more inclusive educational system might not result in changes in PISA outcomes for a few cycles, although such changes could still contribute to students' sense of belonging in their schools.

Providing high-quality education to migrant students and addressing the performance gap are both related to inclusive education and to developing policies related to language as well as social integration and cultural acceptance (Klinger et al. 2018). A central question addresses the extent to which national assessment policies and levels of standardisation regulate assessment at the classroom level. A national curriculum and a national assessment system (e.g. national tests or exams) can be thought of as two key tools in a centralised education system and can be strong drivers to ensuring that all students have access to the same content (Aronson and Laughter 2016; Ayalon and Livneh 2013; Klenowski 2009; Stobart 2005).

Still, these tools are not implemented without risks. For instance, it might be questioned to what extent national tests and exams are developed to accommodate diverse students (Stobart 2005). Previous research has identified major washback effects of national tests and exams (Stobart 2008), limiting what is taught to what is measured in the assessments. However, researchers have found education targeted towards problem-solving and critical thinking to be more effective in raising migrant students' achievement levels compared to more traditionally oriented education, where students are primarily expected to reproduce content and skills they have been taught (Aronson and Laughter 2016; Moschkovich and Nelson-Barber 2009). This finding also indicates that student-centred teaching might be preferred to teacher-centred education. In addition, it should be noted that a national curriculum often represents the majority culture (Klenowski 2009; Stobart 2005); as such, steps must be taken to make the content culturally relevant to migrant students as well.

5.2 Challenges related to teachers' and students' beliefs and conceptions of competence

Several factors have been identified connected to teachers' and students' beliefs and perceptions about teaching, learning and competence in migrant students that might affect these students' possibilities for demonstrating their skills and knowledge in

assessment situations. Prior research has shown that teachers may have stereotyped beliefs about migrant students, for example, believing that they are less able than other students (Moschkovich 2007). Such beliefs are often associated with deficit thinking or the stereotyping of migrant students based on their cultural origin (Portera 2008). Teachers may also have difficulties accepting that students do not act or perform according to the teachers' expectations (Bishop et al. 2003; Fernandes et al. 2017; Rousseau and Tate 2003). This situation can create difficulties for teachers in recognising and acknowledging student competence demonstrated in teaching and assessment situations.

In their studies of New Zealand classrooms, Hunter et al. (2016) observed that participating in inquiry teaching was very challenging for Pāsifika (Pacific Island) students. These students perceived various behaviours, such as approaching other students and challenging their hypotheses, stating clearly that they disagreed with something, and telling peers they may have made a mistake, as being impolite and disrespectful. As a result, the students tended to become passive and retreat from these situations. In general, if a teacher sees this behaviour as a problem with the student and not as a result of the lack of established classroom norms for participation, then the students' understanding of what participation means could be affected; as such, their learning opportunities and their possibilities for demonstrating what they know and can do may also become limited (Hodge and Cobb 2016; Heritage and Wylie 2018). Likewise, students who are from cultures that seek harmony rather than critical reflection (which often requires disagreement or discussion) may have difficulty participating in assessment situations that demand critical thinking and reflection (Civil and Hunter 2015). Prior research has shown that migrant students often hold lower self-beliefs compared to other students and more often display behaviour that may be interpreted as learned helplessness, a situation that further adds to the various factors that tend to keep migrant students back (Moschkovich 2007; Moschkovich and Nelson-Barber 2009). As noted previously, however, such helplessness should instead be interpreted as being reflective of the expected behaviours and values of non-Western cultures. These factors often lead students to withdraw from teaching and assessment situations and, as such, can hinder academic success.

For assessment practices to be culturally responsive and fair, cultural competence and awareness must influence teachers' perceptions of cultural differences and school practices. Teachers must be self-aware and reflective, and they must question their beliefs about students, about instruction, and about the social context of instruction (Hollins 2015). They must also be aware of the composition of their classes and consider which assessment strategies will be suitable for students from particular cultural backgrounds (Qualls 1998). Culturally responsive assessment thus challenges teachers' personal and professional development and calls for a vision of society that is cohesive, just and respectful of differences and which offers genuine opportunities (Portera 2008) for effective and fair EA.

5.3 Challenges related to communication and participation

Padilla (2001) considers several factors that may cause and perpetuate educational barriers that prevent migrant students from taking part in standardised testing, such as

issues related to formal education, their proficiency in the language of instruction in the new country, the length of residence in the new country and their level of acculturation to the dominant culture. These barriers might be outcomes of deficit thinking (Portera 2008), for instance, from viewing migrant students as less knowledgeable or less prepared to participate in the assessment compared to majority students because they do not know the language of instruction well enough (Barwell 2009). Migrant students who are also refugees may be even more vulnerable than other migrant students, as their education has likely been disrupted by the conflicts they have fled from or by what they experienced during the process of moving and resettling (Bilgili et al. 2018).

Several authors, including Özerk and Whitehead (2012), Padilla (2001), Peña et al. (1992) and Portera (2008), pointed to the difficulties in assessing students whose first language is not the dominant language of assessment. Portera (2008) identified several potential pitfalls, including perceiving the students themselves as a problem or considering them at risk and in need of compensatory education (similar to special needs students); another potential pitfall is inaccurately attributing problems to the students, which may prevent the recognition of restricting factors within educational and assessment practices. Özerk and Whitehead (2012) further warned that translating tests from the dominant language into the languages of indigenous cultures presents 'a threat to indigenous ways of knowing' (p. 546; see also Peña et al. 1992). Similar threats might be observed if assessments are translated into the language of migrant students.

Fernandes et al. (2017) challenged the possible deficit perception of bilingual students in mathematics education. They proposed the application of assessment tasks that afford multiple modes of engaging with the tasks; they also proposed the acceptance of multimodal explanations to allow for assessment practices that can be used to better understand what bilingual students know and can do in mathematics. This view might represent a more culturally responsive practice compared to mere translation. Other academics (e.g. Klenowski 2009; Solano-Flores and Nelson-Barber 2001) previously voiced similar views in their calls for culturally valid assessments.

Educational assessment can be viewed as promoting individualism by focusing on assessing individual students (Stobart 2005), which is related both to formative and summative assessment, and by providing individual students with feedback or a grade. However, students may come from backgrounds where collective efforts are more valuable and, as such, may feel more comfortable in situations of collective participation and assessment. According to Stobart (2005), little research has examined the impact of individualism on the assessment of collective-oriented students. Klenowski (2009) asked whose curriculum is assessed, and by which tasks, and she questioned the access that diverse students have to assessment tasks (see also Kirova and Hennig 2013). Is a range of cultural knowledge reflected in what is regarded as a valid response, or might cultural knowledge mediate response patterns to assessment tasks that alter the construct under assessment? (Gipps and Murphy 1994). Because both assessment format and content influence the ways in which migrant students might participate in assessment situations, both factors influence the validity of the assessment, as do the required response formats.

Much previous research has highlighted the role of the teacher in creating CRA-oriented practices for CA (e.g. Heritage and Wylie 2018; Moschkovich and Nelson-Barber 2009). This focus should not be interpreted as CRA being teacher-centred, however. Rather, researchers have highlighted student-centred practices as being

culturally responsive, as student-centred teaching and assessment practices are typically more sensitive to and cognisant of student differences and needs (Aronson and Laughter 2016; Heritage and Wylie 2018; Klenowski 2009).

5.4 Challenges related to student background factors

Students' backgrounds might also affect their performance and possibilities for success in different school systems. Arikan et al. (2017) utilised PISA data to investigate the educational achievement of Turkish migrant students who had migrated to seven European countries by comparing them to mainstream students in the same countries and with mainstream Turkish students in Turkey. The performance differences between Turkish migrant students and other students and differences between groups who had migrated to different European countries were examined by applying individual-level characteristics, such as SES, and country-level characteristics, such as multicultural policies. After controlling for the students' economic, social and cultural status, the authors found medium-level effect differences between migrant and mainstream students. Interestingly, the Turkish migrant students performed better at reading in countries with a high score on the MIPEX index, but when compared to mainstream Turkish students in Turkey, the migrant Turkish students were less successful, which indicates that language may have played a role in the outcomes. Arikan et al. (2017) pointed to the complex role of language in the achievement gap. They argued that Turkish migrant students may be insufficiently exposed to the Turkish language in a rich or sophisticated way in their early years, which would have implications for the students' cognitive development.

Hopson and Hood (2005) proposed the possibility of narrow culture-centric perspectives, techniques and standards that might imbue test development and, as a result, the constructs and formats used to test student learning. Their interpretation is in line with the views of Klenowski (2009) and Stobart (2005, 2008). Such narrow perspectives are not in accordance with CRA principles. Rather, Hopson and Hood (2005) promoted CRA as being contextualised and sensitised to cultural ways of responding. For classroom-level assessment, rather than training students to work with existing assessment formats, implementing changes in teaching approaches used in diverse classrooms may be more valid and efficient, even though doing so can be challenging. Siemon et al. (2004), for instance, prepared teachers to work with students by using rich tasks in mathematics. The teachers were trained to develop assessment formats that would allow minority and Aborigine students to demonstrate their problem-solving competence in mathematics. Although the teachers were sensitised to their students' cultural differences, the attempts failed, in the sense that the minority students did not improve their scores on standardised assessments. Siemon et al. (2004) initiated this professional development research to help teachers align their teaching approaches with the curriculum and with external, large-scale assessment. They, indeed, found that the teaching practices developed by the teachers involved in the study showed similarities to what might be termed 'culturally responsive mathematics education' (cf. Moschkovich and Nelson-Barber 2009). Thus, from a culturally responsive perspective, the project could be considered a success.

6 Towards culturally responsive assessment

As Moschkovich and Nelson-Barber (2009) have noted, Western European and American values have long dominated educational practices and EA. They proposed that teachers need to understand how cognitive approaches and the sociocultural origins of the classroom afford or constrain student participation in assessment. In this case, one might question what is generally considered valid, as what determines the validity of content is cultural.

For CA to be valid, the feedback must move learning forward (Brookhardt 2009; Stobart 2012). In order to determine if teacher feedback is valid and can affect learning, whether the feedback contributes to significant learning outcomes must be scrutinised (Daugherty et al. 2012). School policies, as well as national assessment policies, might also influence teacher practices (Harlen 2012). For instance, external assessments and 'teaching to the test' can disrupt learning and assessment practices (Harlen 2012; Stobart 2005). Solano-Flores and Nelson-Barber (2001) argued that a more refined concept of cultural validity needs to be developed, as sociocultural context influences values, beliefs, expectations, communication patterns, teaching and learning styles and epistemologies inherent in students' cultural background. They proposed that current approaches to manage student diversity in EA are limited and lack a sociocultural perspective. Similar ideas can be found in the work of Stobart (2005), who claimed that fairness is mainly a sociocultural issue, not a technical one.

To practise assessment in diverse classrooms, teachers require insights into their students' cultural ways of expressing themselves (Civil and Hunter 2015; Fernandes et al. 2017; Kirova and Hennig 2013). Hunter et al. (2016), for instance, claimed that language is a key aspect of cultural identity; they linked Pāsifika students' failure in the educational system to structural inequalities caused by the disconnection and dismissal of the students' cultural values, understanding and experiences. As they asserted, Pāsifika students who are fluent in their own language and have a rich background of knowledge and experiences will fail unless they have teachers who establish respectful and reciprocal relations to the students and their parents.

Civil and Hunter (2015) found that non-dominant minority students, who may withdraw from classroom discourse or struggle to express their knowledge in the dominant culture, built relationships when they encountered teachers who supported their students and encouraged them to engage in classroom discussion, although that happened only when the teachers were aware of the students' cultural ways of expressing themselves. Helping teachers and those studying to be teachers to become sensitised to diverse students is no doubt challenging, but a few studies show promising results. In the 'Professional Development School' project, for instance, Wong and Glass (2005) found that teacher students developed more positive attitudes towards students' cultural and linguistic diversity. In addition, teachers and teacher educators who participated in the same project were given an opportunity to develop their own teaching and assessment practices to better accommodate diverse students. Siemon and colleagues' research (Siemon et al. 2004) also showed promising outcomes for cultural diversity.

To address teachers' deficit perceptions of migrant students, Fernandes et al. (2017) suggested that we move from summative approaches to more formative approaches to assess such students. Their proposal also relates to classroom-based assessment practices.

As the findings from the current literature review indicate, assessment strategies that are more culturally fair or culturally responsive to minority students are needed. In the following paragraphs, we discuss how different assessment strategies, including strategies for AfL in diverse schools, can be used to achieve the goal of CRA, which also requires the use of a range of assessment formats and strategies (Aronson and Laughter 2016; Castagno and Brayboy 2008; Espinosa 2005). Prior research (Authors, submitted) identified four strategies with the potential to be culturally responsive: (1) performance-based assessment (cf. Baker, O'Neil and Linn 1993; Hood 1998a; Kim and Zabelina 2015), (2) peer- and self-assessment (Heritage and Wylie 2018; O'Hara et al. 2015), (3) creativity assessment (Kim and Zabelina 2015; Hempel and Sue-Chan 2010) and (4) dynamic assessment (Lidz 2001).

1. In performance-based assessment, the use, application and demonstration of knowledge and skills are evaluated, as opposed to merely the recall of knowledge. Such assessment strategies apply to a wide variety of tasks, with a focus on open-ended tasks, higher-order and complex skills, context-sensitive strategies and complex problems that require several types of performance and significant student time. These assessments may be conducted by individual or group performance and may involve a significant degree of student choice (Baker et al. 1993, p. 1211). As such, performance-based assessment may accommodate both individually and collectively oriented students.

2. Peer- and self-assessment are both integral to AfL (Hayward 2012) and can potentially create more equitable assessment practices in multicultural classrooms (Heritage and Wylie 2018; O'Hara et al. 2015). In peer- and self-assessment, the students are active partners who assess their own performance or that of their peers and share responsibility for assessment with the teacher. To enable such practices, the teacher should establish classroom norms that foster collaboration, trust and the appreciation of differences (Heritage and Wylie 2018). Both peer- and self-assessment can include feedback on evidence from a variety of assessment formats, such as tests, oral presentations, portfolios and skilled behaviour (Topping 2009), or the assessment may be conducted in response to classroom activities, such as problem-solving tasks.

3. In creativity assessment, strategies are used to assess students' abilities for creative thinking. Kim and Zabelina (2015) defined creativity as that which produces 'something that is novel and useful' (p. 136) and showing few differences across gender or ethnicity. Both Hempel and Sue-Chan (2010) and Kim and Zabelina (2015) advocated for the use of creativity assessment to avoid cultural bias.

4. Dynamic assessment strategies are used to consider students' language proficiency. Lidz (2001) described the central feature of dynamic assessment as assessment that relies on active interaction between students and assessors. The assessor observes and records 'events that are presumed to exist within the child' (Lidz 2001, pp. 524–525). This approach focuses on tasks for completion as the teacher analyses 'the child's approach to problem solving and guides the child towards success through promoting mastery of generalizable principles and strategies' (Lidz 2001, p. 526). Student activities are also examined in terms of the 'process demands of the learner (for example attention, perception, memory, conceptual and executive

demands), and the learner is observed and analysed in terms of the application of the process' (Lidz 2001, pp. 526–527).

These four strategies can be used in both formative and summative formats at the classroom level in diverse classrooms. Each can be thought of as a strategy that focuses on the assessment of student competence; as such, they may be perceived as summative assessment strategies. In contrast, formative assessment, such as AfL, focuses on the learning process and how different students utilise their knowledge, skills and strategies in further learning. Applied formatively, the four strategies can be used to elicit information about what students know and can do, which can then be used to provide feedback to the learners. In particular, the above strategies can be used to collect information about migrant students' prior knowledge and their approaches to learning.

In a classroom where day-to-day assessment has a formative or AfL focus, the community of learning, that is, the roles and responsibilities of teachers and students, must be renegotiated (Hayward 2012). Teachers need to listen to student voices, and students should be encouraged to move from being peripheral to being central community members (DeLuca et al. 2018; Kirova and Hennig 2013). AfL necessitates trust between teachers and students (Stobart 2012), and both teachers' and students' assessment output should be treated as valid. Doing so might be challenging for those students who are from cultures where student voices are not traditionally heard, where teaching is teacher-centred, or where students should show due respect for and listen to their elders. Prior research has shown that such activities are challenging for students from some cultures (e.g. Civil and Hunter 2015).

Some might claim that for students to understand teacher feedback, the feedback should be in line with the students' beliefs about teaching and learning and their understanding of the relationship between, and the responsibilities of, students and teachers. An alternative view may be that teaching should help students and teachers together to develop a didactical contract, where students can take on those roles and responsibilities that have often been identified as best practices to support learning (Hodge and Cobb 2016; Hunter et al. 2016; Siemon et al. 2004) and that are in line with AfL practices. Peddar and James (2012), as well as Siemon et al. (2004), argued that professional development is one condition for teachers to adopt AfL. This situation might apply especially if teachers are to adopt AfL in diverse classrooms, as doing so will add to the complexity of their classrooms. A change in the assessment focus might also yield students' understanding of assessment. Nayir et al. (2019) found that although the foundations for CRA are beginning to take shape in some European countries, a significant need still exists for training and professional development in order to sensitise and prepare teachers for CRA practices.

7 Final words

This review examined literature from the complex and varied field of multicultural education and assessment to help us understand how CRA can be achieved in multicultural education systems; the review also examined research on migrant, minority and indigenous students. Although this literature is well suited for identifying the challenges and obstacles that migrant students might face, the different foci of the previous

research studies (e.g. closing the achievement gap, addressing language issues or civil rights) simultaneously make it difficult to tease out what might constitute CRA in a globalising world. In this article, we have interpreted cultural responsiveness to mean respecting, being sensitive to and being cognisant of cultural variations in ways of thinking and knowing, learning, meaning-making, values and beliefs. As such, numerous researchers have proposed ways in which teachers can be sensitised to their students' cultural ways of expressing themselves that might help teachers move towards the use of more equity-oriented and inclusive practices. Although much of the previous research reflects a more static view of culture and how cultural awareness might be utilised in teaching, we argue that having a more fluid concept of culture is necessary (e.g. OECD 2016; Hodge and Cobb 2016) by focusing on the establishment of inclusive classroom norms that allow all students to participate in classroom activities.

To ensure that migrant students have the same opportunities to both learn and demonstrate their competence, we argue that CRA is necessary in both summative and formative assessment formats. Based on our review, we further argue that focusing on AfL in and of itself is insufficient for yielding valid assessments of migrant students. Teachers need to understand students' cultural ways of communicating and participating to provide students with culturally appropriate ways of demonstrating what they know (Hodge and Cobb 2016; Kirova and Hennig 2013; Moschkovich and Nelson-Barber 2009). Teachers and schools also need to be attuned to language issues by providing opportunities for migrant students to learn and use the language of instruction as well as to develop sound conceptual understandings (Gay 2018; Klenowski 2009; Stobart 2008). Furthermore, educators need to understand that multilingual students' use of their own/other language(s) supports learning and mastery of the language of instruction (Barwell 2009).

Establishing classroom norms, or a didactical contract, allows students to become central participants in the local CA community, negotiating and sensitising teachers and students in how to ask and respond to questions (DeLuca et al. 2018; Hodge and Cobb 2016). The establishment of culturally aware classroom norms will aid in the process of eliciting evidence that can support both learning and participation (Kirova and Hennig 2013). We argue that such norms can be best established in inclusive and student-oriented classrooms, where teachers and others recognise diversity as a source of differences in participation and meaning-making, while acknowledging individual students' varying contributions as being valid and interesting. As such, CRA practices become a means to ensure the valid assessment of migrant students. The prior research discussed in this review, however, indicates that teacher and student beliefs might stand in the way of such practices, which means that having transparency in assessment and teaching activities and criteria is crucial. Achieving alignment between curricula, teaching and learning practices and assessment activities is also vital to achieving transparency and CRA. Future research on the assessment of migrant students in diverse classrooms needs to focus not only on how teachers can be sensitised to the culture and language issues discussed in this article but also on how both assessment paradigms and national policies can influence classroom assessment.

References

American Evaluation Association (AEA) (2011). Statement on cultural competence on evaluation. http://www.eval.org/p/cm/ld/fid=92. Accessed 1 Apr 2019.

Arikan, S., van de Vijver, F. J. R., & Yagmur, K. (2017). PISA mathematics and reading performance differences of mainstream European and Turkish immigrant students. *Educational Assessment, Evaluation and Accountability, 29*(3), 229–246. https://doi.org/10.1007/s11092-017-9260-6.

Arnaut, K., Blommaert, J., Rampton, B., & Spotti, M. (2016). Introduction: superdiversity and sociolinguistics. In B. Rampton & J. Blommaert (Eds.), Language and superdiversity (pp. 31–58). London: Routledge.

Aronson, B., & Laughter, J. (2016). The theory and practice of culturally relevant education: a synthesis of research across content areas. *Review of Educational Research, 86*(1), 163–206.

Ayalon, H., & Livneh, I. (2013). Educational standardization and gender differences in mathematics achievement: a comparative study. *Social Science Research, 42*(2), 432–445.

Baker, E. L., O'Neil Jr., H. F., & Linn, R. L. (1993). Policy and validity prospects for performance-based assessment. *American Psychologist, 48*(12), 1210–1218.

Barwell, R. (2009). Summing up: teaching and learning mathematics in a multilingual world. In R. Barwell (Ed.), *Multilingualism in mathematics classrooms: global perspectives* (pp. 161–180). Bristol: Multilingual Matters.

Basterra, M. D. R., Trumbull, E., & Solano-Flores, G. (2011). *Cultural validity in assessment: addressing linguistic and cultural diversity.* New York: Routledge.

Bernstein, B. (2000). *Pedagogy symbolic control and identity: theory, research, critique* (rev. ed.). Lanham: Rowman & Littlefield.

Bilgili, O., Volange, L., & Klinger, D. (2018). Immigrant student achievement and the performance disadvantage. In L. Volante, D. Klinger, & O. Bilgili (Eds.), *Immigrant student achievement and educational policy* (pp. 3–15). Cham: Springer International Publishing.

Bishop, R., Berryman, M., Tiakiwai, S., & Richardson, C. (2003). *Te Kotahitanga: the experience of year 9 and 10 Māori students in mainstream classrooms.* Wellington: Ministry of Education.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102.

Black, P., & Wiliam, D. (2012a). Assessment for learning in the classroom. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 11–32). London: Sage.

Black, P., & Wiliam, D. (2012b). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 206–229). London: Sage.

Boesen, J. (2006). Assessing mathematical creativity: Comparing national and teacher-made tests, explaining differences and examining impact. Doctoral thesis. Umeå University, Faculty of Science and Technology, Department of Mathematics and Mathematical Statistics.

Bourdieu, P. (1984). *Distinction.* London: Routledge & Keegan Paul.

Bourdieu, P. (1990). Reproduction in education, society and culture. London: Sage.

Bradshaw, C. P., O'Brennan, L. M., & McNeely, C. A. (2008). Core competencies and the prevention of school failure and early school leaving. New directions for child and adolescent development, 122, 19–32. DOI: https://doi.org/10.1002/cd.226.

Brentnall, R., & Hodge, A. (1984). *Policies on multicultural education in Australia: an overview.* Sydney: Education Commission of New South Wales, Sydney College of Advanced Education, Multicultural Centre.

Brookhardt, S. M. (2009). Assessment and examinations. In L. J. Saha & A. G. Dworkin (Eds.), *International handbook of research on teacher and teaching* (pp. 723–738). Boston: Springer.

Brown-Jeffy, S., & Cooper, J. E. (2011). Toward a conceptual framework of culturally relevant pedagogy: an overview of the conceptual and theoretical literature. *Teacher Education Quarterly, 38*(1), 65–84.

Burkhardt, H., & Schoenfeld, A. (2018). Assessment in the service of learning: challenges and opportunities or plus ça change, plus c'est la même chose. *ZDM, 50*(4), 571–585. https://doi.org/10.1007/s11858-018-0937-1.

Cahill, D. (1986). An evaluation of Australia's multicultural education program. *Journal of Multilingual and Multicultural Development, 7*(1), 55–69.

Castagno, A. E., & Brayboy, B. M. J. (2008). Culturally responsive schooling of indigenous youth: a review of the literature. Review of Educational Research, 78(4), 941–993.

Civil, M., & Hunter, R. (2015). Participation of non-dominant students in argumentation in the mathematics classroom. *Intercultural Education, 26*(4), 296–312.

Cook, W. W. (1951). The functions of measurement in the facilitation of learning. In E. F. Lindquist (Ed.), *Educational measurement* (pp. 3–46). Washington, DC: American Council on Education.

Crul, M., & Schneider, J. (2010). Comparative integration context theory: participation and belonging in new diverse European cities. *Journal of Ethnic and Racial Studies, 33*(7), 1249–1268. https://doi.org/10.1080/01419871003624068.

Crul, M., Schnell, P., Herzog-Punzenberger, B., Wilmes, M., Slootman, M., & Aparicio-Gomez, R. (2012). School careers of second-generation youth in Europe: which education systems provide the best chances for success? In M. Crul, J. Schneider, & F. Lelie (Eds.), *The European second generation compared: Does the integration context matter?* (pp. 101–164). Amsterdam: Amsterdam University Press.

Cumming, J. J., & van der Kleij, F. M. (2016). Effective enactment of assessment for learning and student diversity in Australia. In D. Laveault & L. Allal (Eds.), *Assessment for learning: meeting the challenge of implementation* (pp. 55–73). Cham: Springer International Publishing.

Cummins, J. (2015). Intercultural education and academic achievement: a framework for school-based policies in multilingual schools. *Intercultural Education, 26*(6), 455–468. https://doi.org/10.1080/14675986.2015.1103539.

Daugherty, R., Black, P., Ecclestone, K., James, M., & Newton, P. E. (2012). Alternative perspectives on learning outcomes: challenges for assessment. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 72–86). London: Sage.

DeLuca, C., Chapman-Chin, A. E. A., LaPointe-McEwan, D., & Klinger, D. A. (2018). Student perspectives on assessment for learning. *Curriculum Journal, 29*(1), 77–94. https://doi.org/10.1080/09585176.2017.1401550.

Devine, D., Kenny, M., & Macneela, E. (2008). Naming the 'other': children's construction and experience of racisms in Irish primary schools. *Race Ethnicity and Education, 11*(4), 369–385. https://doi.org/10.1080/13613320802478879.

Eder, F., Neuweg, G. H., & Thonhauser, J. (2009). Leistungsfeststellung und leistungsbeurteilung [Performance assessment and performance appraisal]. In W. Specht (Ed.), *Nationaler Bildungsbericht Österreich 2009, band 2: Fokussierte analysen bildungspolitischer schwerpunktthemen [National Education Report Austria 2009, volume 2: Focused analyses of key educational topics]* (pp. 245–267). Graz: Leykam.

Espinosa, L. M. (2005). Curriculum and assessment considerations for young children from culturally, linguistically, and economically diverse backgrounds. *Psychology in Schools, 42*(8), 837–853.

Fernandes, A., Kahn, L. H., & Civil, M. (2017). A closer look at bilingual students' use of multimodality in the context of an area comparison problem from a large-scale assessment. *Educational Studies in Mathematics, 95*(3), 263–282.

Ford, D. Y., & Kea, D. (2009). Creating culturally responsive instruction: for students' sake and teachers' sake. *Focus on Exceptional Children, 41*, 1–18.

Frierson, H. T., Hood, S., Hughes, G. B., & Thomas, V. G. (2010). A guide to conducting culturally responsive evaluations. In J. Frechtling (Ed.), *The 2010 user-friendly handbook for project evaluation* (pp. 75–96). National Science Foundation, Directorate for Education and Human Resources. https://www.purdue.edu/research/docs/pdf/2010NSFuser-friendlyhandbookforprojectevaluation.pdf. Accessed 29 Oct 2019.

Gardner, J. (2012). Assessment and learning: introduction. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 1–10). London: Sage.

Gay, G. (2018). *Culturally responsive teaching: Theory, research and practice* (3rd ed.). New York: Teachers College Press, Columbia University.

Gipps, C., & Murphy, P. (1994). A fair test? Assessment, achievement and equity. Paper presented at the American Educational Research Association conference, New Orleans, LA.

Grant, M., & Booth, A. (2009). A typology of 14 review types and associated methods. *Health Information and Libraries Journal, 26*(2), 91–108.

Harlen, W. (2012). On the relationship between assessment for formative and summative purposes. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 87–102). London: Sage.

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

Hattie, J. A. C., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Hayward, L. (2012). Assessment and learning: the learner's perspective. In J. Gardner (Ed.), Assessment and learning (2nd, pp. 125–139). London: Sage.

Hempel, P. S., & Sue-Chan, C. (2010). Culture and the assessment of creativity. *Management and Organization Review, 6*(3), 415–435. https://doi.org/10.1111/j.1740-8784.2010.00189.

Heritage, M., & Wylie, C. (2018). Reaping the benefits of assessment for learning: achievement, identity, and equity. ZDM, 50(4), 729–741.

Herzog-Punzenberger, B., Le Pichon-Vorstman, E., & Siarova, H. (2017). *Multilingual education in the light of diversity: lessons learned. NESET II report.* Luxembourg: Publications Office of the European Union. https://doi.org/10.2766/71255.

Hodge, L. L., & Cobb, P. (2016). Two views of culture and their implications for mathematics teaching and learning. *Urban Education*. https://doi.org/10.1177/0042085916641173.

Hollins, E. R. (2015). *Culture in school learning: revealing the deep meaning* (3rd ed.). New York: Routledge.

Hood, S. (1998a). Introduction and overview. Assessment in the context of culture and pedagogy: a collaborative effort, a meaningful goal. Journal of Negro Education, 67(3), 184–186. DOI: https://doi.org/10.2307/2668187.

Hood, S. (1998b). Culturally responsive performance-based assessment: conceptual and psychometric considerations. *Journal of Negro Education, 67*(3), 187–196. https://doi.org/10.2307/2668188.

Hopson, R., & Hood, S. (2005). An untold story in evaluation roots: Reid E. Jackson and his contributions toward culturally responsive evaluation. In S. Hood, R. Hopson, & H. Frierson (Eds.), The role of culture and cultural context (pp. 87–104)*.* Greenwich: Information Age Publishing.

Hunter, J., Hunter, R., Bills, T., & Cheung, I. (2016). Developing equity for Pāsifika learners within a New Zealand context: attending to culture and values. New Zealand Journal for Educational Studies, 51(2), 197–209.

Kim, K. H., & Zabelina, D. (2015). Cultural bias in assessment: can creativity assessment help? *International Journal of Critical Pedagogy, 6*(2), 129–147.

Kirova, A., & Hennig, K. (2013). Culturally responsive assessment practices: examples from an intercultural multilingual early learning program for newcomer children. *Power and Education, 5*(2), 106–119. https://doi.org/10.2304/power.2013.5.2.106.

Klenowski, V. (2009). Australian indigenous students: addressing equity issues in assessment. *Teaching Education, 20*(1), 77–93.

Klinger, D., Volante, L., & Bilgili, O. (2018). Cross-cultural approaches to mitigating the immigrant student performance disadvantage. In L. Volante, D. Klinger, & O. Bilgili (Eds.), *Immigrant student achievement and educational policy* (pp. 197–206). Cham: Springer International Publishing.

Köller, O. (2005). Formative assessment in classrooms: a review of the empirical German literature. In J. Looney (Ed.), *Formative assessment: Improving learning in secondary classrooms* (pp. 265–279). Paris: Organisation for Economic Co-Operation and Development (OECD).

Leahy, S., & Wiliam, D. (2012). From teachers to schools: Scaling up professional development for formative assessment. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 49–71). London: Sage.

Lemu, D. (2015). Academic achievement gap: the case of children of immigrants. https://sophia.stkate.edu/msw_papers/481. Accessed 29 Oct 2019.

Lidz, C. S. (2001). Multicultural issues and dynamic assessment. In L. A. Suzuki, J. G. Ponterotto, & P. J. Meller (Eds.), *Handbook of multicultural assessment* (2nd ed., pp. 523–539). San Francisco: Jossey-Bass.

Lingard, R. (1983). Multicultural education in Queensland: the assimilation of an ideal. *Discourse, 4*(1), 13–31.

Mazurek, K., & Kach, N. (1983). Culture and power: educational ideologies in multicultural Canada. *New Education, 5*(2), 47–59.

Meissner, F., & Vertovec, S. (2015). Comparing super-diversity. *Ethnic and Racial Studies, 38*(4), 541–555. https://doi.org/10.1080/01419870.2015.980295.

Migrant Integration Policy Index (MIPEX) (2015). http://www.mipex.eu/. Accessed 29 Oct 2019.

Moschkovich, J. N. (2007). Beyond words to mathematical content: assessing English learners in the mathematics classroom. In A. Schoenfeld (Ed.), *Assessing mathematical proficiency* (pp. 345–352). New York: Cambridge University Press.

Moschkovich, J. N., & Nelson-Barber, S. (2009). What mathematics teachers need to know about culture and language. In B. Greer, S. Mukhopadhyay, S. Nelson-Barber, & A. Powell (Eds.), *Culturally responsive mathematics education* (pp. 111–136). New York: Routledge; Taylor & Francis Group.

Murphy, P. (1986). Multicultural issues and the role of educational administrators. Education and Society, 4(2), 43–50.

National Research Council (NRC) (2001). Knowing what students know: the science and design of educational assessment. Committee on the Foundations of Assessment. Pelligrino, J., Chudowsky, N., & Glaser, R. (Eds.), Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Nayir, F., Brown, M. D., Burns, D., O'Hara, J., McNamara, G., Nortvedt, G. A., Skedsmo, G., Gloppen, S. K., & Wiese, E. F. (2019). Assessment with and for Migration Background Students-Cases from Europe. *Eurasian Journal of Educational Research*. (79), s 39–69.

Nortvedt, G. A., Santos, L., & Pinto, J. (2016). Assessment for learning in Norway and Portugal: the case of primary school mathematics teaching. *Assessment in Education: Principles, Policy & Practice, 23*(3), 377–395. https://doi.org/10.1080/0969594X.2015.1108900.

O'Hara, J., McNamara, G., & Harrison, K. (2015). Culture changes, Irish evaluation and assessment traditions stay the same? Exploring peer- and self-assessment as a means of empowering ethnic minority students. In S. Hood, R. Hopson, & H. Frierson (Eds.), *Continuing the journey to reposition culture and cultural context in evaluation theory and practice* (pp. 205–231). Charlotte: Information Age Publishing.

OECD. (2015). *Helping migrant students to succeed at school and beyond*. Paris: OECD https://www.oecd.org/education/Helping-immigrant-students-to-succeed-at-school-and-beyond.pdf.

OECD. (2016). *Global competency for an inclusive world*. Paris: OECD https://www.oecd.org/education/Global-competency-for-an-inclusive-world.pdf. Accessed 29 Oct 2019.

Özerk, K., & Whitehead, D. (2012). The impact of national standards assessments in New Zealand and national testing protocols in Norway on indigenous schooling. *International Electronic Journal of Elementary Education, 4*(3), 545–561.

Padilla, A. M. (2001). Issues in culturally appropriate assessment. In L. A. Suzuki, J. G. Ponterotto, & P. J. Meller (Eds.), *Handbook of multicultural assessment* (2nd ed., pp. 5–27). San Francisco: Jossey-Bass.

Peddar, D., & James, M. (2012). Professional learning as a condition for assessment for learning. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 33–48). London: Sage.

Peña, E., Quinn, R., & Iglesias, A. (1992). The application of dynamic methods to language assessment: a nonbiased procedure. *Journal of Special Education, 26*(3), 269–280. https://doi.org/10.1177/002246699202600304.

Pollitt, A., Marriott, C., & Ahmed, A. (2000). Language, contextual and cultural constraints on examination performance. Paper presented at the 26th Association for Educational Assessment (IAEA) conference, Jerusalem, 14–19 May.

Portera, A. (2008). Intercultural education in Europe: epistemological and semantic aspects. Intercultural Education, 19(6), 481–491. DOI: https://doi.org/10.1080/14675980802568277.

Qualls, A. L. (1998). Culturally responsive assessment: development strategies and validity issues. *Journal of Negro Education, 67*(3), 296–301. https://doi.org/10.2307/2668197.

Raines, T. C., Dever, B. V., Kamphaus, R. W., & Roach, A. T. (2012). Disproportionate placement in special education. *Journal of Negro Education, 81*(3), 283–296. https://doi.org/10.7709/jnegroeducation.81.3.0283.

Reagan, T. (1984). Critical multiculturalism, border knowledge, and the canon: implications for general education and academy. *Journal of General Education, 2*(1), 111–113.

Rousseau, C., & Tate, W. F. (2003). No time like the present: reflecting on equity in school mathematics. *Theory Into Practice, 42*(3), 210–216.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144. https://doi.org/10.1007/BF00117714.

Schmidinger, E., Hofmann, F., & Stern, T. (2015). Leistungsbeurteilung unter berücksichtigung ihrer formativen funktion [Performance assessment taking into account their formative function]. In M. Bruneforth, F. Eder, K. Krainer, C. Schreiner, A. Seel, & C. Spiel (Eds.), *Nationaler Bildungsbericht Österreich 2015: Band 2. Fokussierte analysen bildungspolitischer schwerpunktthemen [National*

Education Report Austria 2015: Volume 2. Focused analyses of key educational topics] (pp. 49–94). Graz, Austria: Leykam.

Schnell, P. (2014). Educational mobility of second-generation Turks: cross-national perspectives. Amsterdam: Amsterdam University Press.

Shepard, L. (2006). Classroom assessment. In R. L. Brennan (Ed.), *Educational measurement* (pp. 623–646). Westport: Greenwood Publishing.

Siemon, D., Enilane, F., & McCarthy, J. (2004). Supporting indigenous students' achievement in numeracy. *Australian Primary Mathematics Classroom, 9*(4), 50–53.

Slee, J. (2010). A systemic approach to culturally responsive assessment practices and evaluation. *Higher Education Quarterly, 64*, 246–260. https://doi.org/10.1111/j.1468-2273.2010.00464.x.

Smyth, E., Darmody, M., McGinnity, F., & Byrne, D. (2009). Adapting to diversity: Irish schools and newcomer students. Economic and Social Research Institute (ESRI) research series, no. 8. Dublin: ESRI.

Solano-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching, 38*(5), 553–573.

Stephens, D. (2007). Culture in education and development principles, practice and policy. Oxford: Symposium Books.

Stobart, G. (2005). Fairness in multicultural assessment systems. *Assessment in Education, 12*(3), 275–287.

Stobart, G. (2008). Testing times: the uses and abuses of assessment. Abingdon: Routledge.

Stobart, G. (2012). Validity in formative assessment. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 233–242). London: Sage.

Topping, K. J. (2009). Peer assessment. *Theory Into Practice, 48*(1), 20–27.

UNESCO (1994). The Salamanca statement and framework for action on special needs education. http://www.unesco.org/education/pdf/SALAMA_E.PDF. Accessed 29 Oct 2019.

Vertovec, S. (2007). Super-diversity and its implications. *Ethnic and Racial Studies, 30*(6), 1024–1054. https://doi.org/10.1080/01419870701599465.

Vertovec, S. (2010). Towards post-multiculturalism? Changing communities, conditions and contexts of diversity. *International Social Science Journal, 61*(199), 83–95. https://doi.org/10.1111/j.1468-2451.2010.01749.x.

Wiliam, D. (2007). Keeping learning on track. In F. K. J. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1053–1098). Charlotte: Information Age Publishing.

Wong, P., & Glass, R. (2005). Assessing a Professional development School approach to preparing teachers for urban schools serving low-income, culturally and linguistically diverse communities. *Teacher Education Quarterly, 32*(3), 63–77