Language Test Misuse

Cecilie Hamnes Carlsen & Lorenzo Rocca

Published online: 05 Aug 2021.

Submit your article to this journal ↗

Article views: 113

View related articles ↗

View Crossmark data ↗

Routledge
Taylor & Francis Group

Language Test Misuse

Cecilie Hamnes Carlsen 🆔[a] and Lorenzo Rocca 🆔[b]

aWestern Norway University of Applied Sciences, Bergen, Norway; bSocietà Dante Alighieri, Rome, Italy

ABSTRACT

During the past two decades, an increasing number of European countries have introduced language requirements for residency, citizenship, and sometimes even for entry to the country and family reunification. As a result, democratic rights as well as basic human rights have come to depend upon an individual's ability to obtain a certain score on a language test and the consequences of failing may be detrimental. In the field of language testing, this use of language tests is often referred to as *test misuse*, yet the term lacks a precise definition in the literature. In this paper we propose a definition of test misuse in relation to language tests for migration purposes and focus particular attention on low-literate adult migrants for whom the requirements pose a considerable barrier. The main purpose of this paper is to address the question why language tests are being misused in migration policies, exploring linguistic, political as well as test theoretical explanations. We suggest that a more central role of test misuse in validity theory is essential in order to remedy its lack of research focus in our field.

Introduction: What is language test misuse?

During the past two decades, an increasing number of European countries have introduced language requirements as part of their migration policy making passing language tests conditional for residency, citizenship, and sometimes even for entry to the country and family reunification (Bruzos, Erdocia, & Khan, 2018; Rocca, Deygers, & Carlsen, 2020; Slade, 2010).[1] As a result, democratic rights as well as basic human rights have come to depend on an individual's ability to obtain a certain score on a language test, something which is not equally easy for all. With this new trend in migration policies, language tests typically developed to measure communicative language ability, are being interpreted and used as if they measure something else, such as the willingness or ability to successfully integrate (Van Avermaet & Gysen, 2009, p. 119). In the professional field of language testing and assessment, this use of language tests is often referred to as *test misuse*. The term, however, is underdefined in the literature and appears to entail at least two somewhat different contents: On the one hand, test misuse refers to the use of a test for a purpose other than the one for which it was originally designed (Fulcher & Davidson, 2009). A similar conception of test misuse as a mismatch between the intended and the actual use, is also central in Davis (2012), and Wall (2012). An alternative characterization of test misuse

---

[1]Australia and the USA have a long tradition for using language tests as part of immigration or citizenship policies (Kunnan, 2009; McNamara & Roever, 2006; Shohamy & Menken, 2015).

focuses on the negative consequences of a test on test takers, regardless of whether these consequences are intentional or not (Shohamy, 1998, 2001, 2009; Van Avermaet & Pulinx, 2014).

In this paper, we propose a definition of test misuse that includes both aspects mentioned above: the use of a test beyond its original purpose and/or negative or harmful consequences of test (score) interpretation and use, irrespective of intentionality. Thus, either one or both of the following would qualify as language test misuse:

- a test that was created to measure a certain *construct* (language) and for a certain purpose, but the scores of which are now being interpreted and used as if they measure something else
- a test that has non-beneficent or harmful *consequences* for test-takers, regardless of whether those consequences were intentional or not

In the context of migration policies, an example of test misuse would be when language tests serve as a proxy for migrants' willingness or ability to successfully integrate, driven by an intentional, though not always explicitly formulated, purpose to exclude certain groups of migrants from entrance, residency or citizenship in order to achieve supposed social cohesion, national security or economic prosperity. In addition, even if a test was introduced to promote language learning and integration, if the consequences for all or certain groups of migrants was to be negative, this would be considered test misuse according to our definition.

The main purpose of this paper is to address the question *why language tests are being misused in migration policies*, in other words to explore possible reasons why language tests are particularly susceptible to political misuse, taking a closer look at linguistic, political as well as test theoretical explanations. We address the question why, given the very widespread misuse of language tests in migration policy today, this question has not received more focus in language test research and literature, proposing as a likely explanation the marginalized status of deliberate test misuse in validity theory. When language tests are being misused in the ways described above, adult migrants and refugees with limited prior schooling and low levels of literacy (hereafter LESLLA learners)[2] are particularly prone to discrimination (Oers, 2020): Not only are LESLLA learners' chances to pass the tests lower, the consequences of failing are also typically very severe since passing the test not only influences their opportunities for education, labour and access to welfare goods, but may also directly impact the chances for a secure future, family reunification, and protection against deportation. A recent study of language requirements and learning opportunities in Europe (Rocca et al., 2020) reveals that this particularly vulnerable learner group is rarely catered for in language learning and language testing policies and few countries have exemptions from requirements for this group. When discussing test misuse, LESLLA learners deserve special attentions since they are likely to be the ones most severely affected by it. This paper therefore starts with a focus on this understudied population in second language acquisition research and language test research alike. Towards the end of the paper, we discuss a possible way forward in order to prevent test misuse in the migration

---

[2]LESLLA is an acronym for Literacy Education and Second Language Learning for Adults (www.lessla.org)

context, emphasizing the important role of test developers in the area of language test advocacy.

Test misuse and low literate learners

Throughout world history, war, conflict, poverty and oppression have driven people to leave their homes to search for safety and better opportunities elsewhere. These same factors are also important barriers for the spread of literacy skills, and explanations as to why, at a global level, more than 770 million adults have no or only very basic literacy skills (UNESCO, 2017). As a consequence, a certain percentage of the refugee population in Western countries are non-literate or low literate with limited prior schooling[3] (see Hooft et al., this issue).

When passing a formal language test[4] is made a condition for residency and citizenship, non-literate and low literate test takers risk being systematically discriminated against, since formal learning and passing tests is more challenging for them than it is for literate learners with more education: Research has shown that LESLLA learners have specific challenges when it comes to learning a second language, their learning progress has been found to be slower and their L2-learning outcome lower than that of higher educated learners (Kurvers, Van De Craats, & van Hout, 2015). They benefit less from language courses, perform less well on tests in general and language tests in particular (Kim, Yoon, Kim, & Kim, 2014; Carlsen, 2017; see also Ruesseler et al., this volume).[5] Their lack of success in language tests is not only caused by a lack in the skills tested, but also by lack of experience with the testing situation and lack of familiarity with test formats commonly used in language tests (Allemano, 2013). Due to a lack of research interest in this learner group (Andringa & Godfroid, 2019), we know significantly less about the second language learning processes and learning needs of LESLLA learners than we do about the learning process of more educated learners (Tarone, 2010). Therefore, second language acquisition theories and empirically based knowledge may only be partly relevant to describe and understand how LESLLA learners acquire a new language and how to best cater for their specific learning needs (Van De Craats, Kurvers, & Young-Scholten, 2006). Teacher training, teaching materials, and teaching methods are also only to a limited degree developed for LESLLA learners.[6]

In 2018, the Council of Europe (CoE) in collaboration with the Association of Language Testers in Europe (ALTE) conducted the fourth survey of language and knowledge of society (KoS) requirements for migrants in CoE member states. As the survey report

---

[3]Important to note, factors disfavouring language learning may cluster. Refugees from zones of conflict may lack schooling and therefore also lack literacy skills, many have an L1 typologically distant from the target language, many refugees suffer from traumas and long-term PTSD, many fear the future not knowing if they will obtain residency/citizenship, in addition to marginalisation, racism and discrimination in social life, education, labour, housing etc.

[4]More often than not, language tests are introduced together with knowledge of society tests, which, since they are normally in the language of the host country, are implicit language tests (Rocca et al., 2020).

[5]Important to acknowledge, however, many LESLLA learners hold other important resources of importance to themselves and to their family, neighbours and friends. These adults should not, only because of their particular challenges in a formal teaching and testing context, be regarded as generally powerless, vulnerable or deficient in relevant social capital (Norton, 2013; Yosso, 2005).

[6]It is also relevant to mention that language and knowledge of society (KoS) test are rarely tailored to LESLLA learners. It is not the focus of this paper how to develop language tests that take LESLLA learners into account, for this, see for instance Carlsen, 2017.

shows (Rocca et al., 2020) only seven out of 40 responding member states did *not* have formal language requirements for migration purposes in 2018 (entrance, residency and/or citizenship), revealing a considerable retrenchment since the first CoE survey was carried out in 2007.

The results show that LESLLA learners seem largely neglected by policymakers when language and KoS-requirements are set in migration policy. This group is seldom taken into account in test development, and policy is rarely based on needs analyses or studies of test impact on this vulnerable learner group. Moreover, despite the fact that LESLLA learners tend to perform better in oral skills than in written skills (Carlsen, 2017), few CoE states set differentiated language requirements in different skills taking this into account. This is somewhat surprising, given that a profiled approach is strongly recommended in the *Common European Framework of Reference* (Council of Europe, 2001) upon which most CoE states base their language requirements. The only countries setting lower requirements in reading/writing than in listening/speaking for one or more contexts are Germany, Italy, Luxembourg, Norway, and the UK.[7] Only ten percent of the countries with requirements have exemptions for LESLLA learners.

The survey also investigated learning opportunities, and found that in this area too, a concern for LESLLA learners is rare: Only half of the responding countries affirmed to have courses targeted to non-literate and low literate learners, and even among these, the number of hours of tuition is rarely sufficient: LESLLA learners need more time to acquire a language, both orally and in writing, yet the number of hours they receive is insufficient to meet their special needs for slower pace and more hours of instruction and to compensate for their literacy gap: Most countries do not offer more than 250 hours of instruction.[8]

Why are language tests misused in migration policy?

Central to the understanding of test misuse in migration policies, is a recognition of the multiple roles of language (Shohamy, 2009). Language is a means of communication, but also a signal of identity and group belonging. One reason why language tests and not tests of mathematics, for example, are prone to political misuse is to be found in this double role of language itself. A second reason for test misuse is to be found in the overt and covert policy arguments used to defend this practice. The double role of language allows a political rhetoric claiming that language requirements are introduced to foster linguistic integration, while a real agenda of exclusion and control, may be hidden. Given the widespread political use of language tests in Europe today, it is surprising that this phenomenon has not received more research interest in the professional field of language testing and assessment. A final reason for test misuse proposed in this paper, then, is the marginalization of deliberate test misuse in validity theory. These reasons will be discussed in more detail below.

---

[7]UK; for pre-entrance requirements, Italy; for temporary residency, Germany; for permanent residency, Luxembourg for citizenship, and Norway for permanent residency and citizenship.

[8]These findings underscore the importance of the recommendations of the CoE in various documents that language requirements for basic rights like entry, residency and citizenship should not exceed A1 in writing and A2 in speaking (Council of Europe, 2014).

Linguistic reasons for language test misuse

In modern language pedagogy, the conception of language as the ability to communicate efficiently and adequately in context is dominating (Celce-Murcia, 2008). This view, known as *communicative competence*, introduced by Hymes in 1972, and further developed for second language learning by Canale & Swain (1980) and for assessment purposes by Bachman (1990) is still the leading theoretical paradigm for language teaching and testing today. In Europe, the central position of communicative teaching and testing has been further promoted through the widespread adoption of the CEFR and the recently published CEFR Companion Volume (Council of Europe, 2020).[9] The explicit construct of most standardized language tests in Europe today, is then, communicative competence (Fulcher, 2000; ALTE, 2020, p. 20).

To fully understand how language and language tests can be misused as gatekeepers in migration policy, it is pivotal to recognize that in addition to its communication-functional role, language has a significant *symbolic role* (Evans, 2019). Within a community of shared sociolinguistic norms, language variation carries social value, which enables others to make assumptions about our identity and group affiliation. Language may reveal who we are and who we want to be. Language is a marker of social and cultural identity, and language clues may signal identity traits like age, gender, sexual orientation, education, occupation, geographic origin, political preferences, and social class (Edwards, 2009). Language is a marker of *group belonging*, and one of group distancing, and as such, language can be a sign through which we distinguish *us* from *them* (Anderson, 2016; Slade & Möllering, 2010) and friends from foes (McNamara, 2020; McNamara & Roever, 2006).

In the context of migration, the ability to speak the national language(s) has become a symbol of *national belonging* (Anderson, 2016; Blackledge, 2009; Extra & Spotti, 2009, p. 126; Oers, Ersbøll, & Kostakopoulou, 2010). The national language is "viewed in ideological terms as part of a national identity embedded with notions that language is an indicator of loyalty, patriotism, belonging, inclusion, and membership" (Shohamy, 2006, p. 174). That a state should be unified by a common language ("one language, one nation, one state"), originally a founding idea in the eighteenth and nineteenth-century nationalism, has been revitalized in Western countries with increased migration since the 1970s and intensified with the post-2001 terrorist fear (Blackledge, 2009; Joppke, 2010, p. 61; Slade, 2010). The inability to speak the national language has become embedded with the social value of non-belonging and may even be interpreted as an *unwillingness to integrate* (Van Avermaet & Gysen, 2009, p. 119). For low-literate migrants and refugees with limited prior schooling and test experience, such ideas are particularly harmful since passing a language test may represent an unsurmountable barrier to them, regardless of their motivation or ambitions (Oers, 2020). This double role of language is what is in play when those in power use language and language tests to detect and keep out those without (McNamara, 2005; Shohamy, 2001). Therefore, as Goodman (2017, p. 237) importantly points out, even though the language requirements for migration purposes are per se symbolic, the consequences for migrants are very real and concrete indeed.

---

[9]The CEFR and the Companion Volume describe communicative language proficiency from the very basic pre-A1 level to the very advanced C2-level. Important to stress, the levels from B2 and above are characterized by a breadth and complexity of language structures, vocabulary and pragmatic control which is not generally defining for native speakers without academic education (Hulstijn, 2015).

The construct of language tests for migration and citizenship purposes can best be understood in terms of the symbolic function of language, a point also made by McNamara who sees it as a "[. . .] pretext for a deeper assessment, of (external) conformity to a national ideology" (McNamara, 2010, p. 19). The explicit construct of language tests for migration purposes, communicative proficiency in the target language, disguises the implicit construct, language as a sign of migrants' willingness, ability or success in the integration process, as underscored by Van Avermaet and Gysen (2009). While it is easy to see the connection between communicative language proficiency and language test performance for different jobs and higher education where functional communication abilities have a real purpose, it is virtually impossible to establish a similar link between language proficiency and language tests for migration purposes: Indeed, it might be argued, it is not even possible to carry out a needs-analysis of the target language use domain for migration tests to come up with domain-relevant tasks or even to define a suitable proficiency level when tests are used to control access to entrance, settlement and citizenship, for, what level of language proficiency is necessary and sufficient for a refugee to be reunited with his or her family? How well does someone need to master the majority language to be granted permanent residency or obtain citizenship?

Political reasons for language test misuse

When the Norwegian coalition government proposed to introduce language requirements for residency and citizenship as a response to the refugee crisis in 2015, the explicitly stated purpose was to "make it less attractive to apply for asylum in Norway" (regjeringen, 2015). Oers et al. (2010, p. 312) find that the intention of controlling migration was explicitly mentioned as a reason for introducing requirements also in Austria, Denmark, France and the Netherlands. More commonly, however, policy makers legitimise the introduction of language tests for entry, residency and citizenship claiming to promote integration (Blackledge, 2009; Joppke, 2010; Strik, Böcker, Luiten, & Oers, 2010, p. 107; Goodman & Wright, 2015).[10] Social scientists and language test researchers alike have questioned whether promoting integration are indeed the real intentions for introducing language requirements for migration purposes, pointing to an implicit intention, or hidden agenda, of exclusion and control (McNamara & Roever, 2006; Shohamy, 2006; Mackenzie, 2010; Slade, 2010; Goodman & Wright, 2015). There are at least three reasons why this is a likely assumption.

Firstly, language requirements are typically introduced as part of general restrictions in migration policies promoted primarily by right-wing parties with an often-explicit aim to keep the number of new migrants and citizens down (Goodman & Wright, 2015; Joppke, 2010). The real purpose, according to political scientists like Goodman, Slade and Joppke, is to signal to voters that one leads a strict immigration policy and are in control of the migrant situation (Slade, 2010, p. 8). The introduction of language requirements for citizenship reflects the policy makers' view of citizenship: Liberal policies in which citizenship is considered a driver for integration (e.g. Sweden), typically have no or only low requirements, while countries with a restrictive immigration policy in which citizenship is

---

[10]The CEFR and the Companion Volume describe communicative language proficiency from the very basic pre-A1 level to the very advanced C2-level.

regarded a privilege or a prize for successful integration (e.g. Denmark), typically introduce strict requirements (Blackledge, 2009; Goodman & Wright, 2015; Joppke, 2010; Midtbøen, 2015; Oers, 2020).

Secondly, comparisons of actual language requirements for entry, residency and citizenship across European countries reveal a considerable degree of divergence in terms of levels of proficiency required for the same policy context (Rocca et al., 2020; Van Avermaet & Pulinx, 2014). Requirements for citizenship, for example, range from no requirement (Sweden and Ireland) to an academic level (B2) (Denmark and Austria).[11] If the communicative language needs of migrants were indeed the real reason for introducing such requirements, one would expect a greater degree of agreement as to what is considered necessary and sufficient for the same context across countries, as argued by McNamara and Shohamy (2008, p. 92), and as apparent when CEFR-levels for admission to higher education was compared across Europe (Deygers, et al., 2018). The fact that there are considerable differences across Europe in terms of size, population, language and social order cannot explain why in some countries immigrants would need academic language skills (B2) to be a citizen while in others, no or only very limited communicative skills would be sufficient for the same context.

A third reason that casts doubt on the argument that migration tests are introduced to support migrants' integration and language learning, is that research on the consequences of such requirements clearly suggests the opposite effect (Pochon-Berger & Lenz, 2014, p. 30): The largest study to date, the INTEC-project investigating the quantitative and qualitative effect of migration tests in nine European countries, finds few (if any) beneficial effects on those subjected to such tests (Strik et al., 2010). Oers (2014) comparing the effect of citizenship requirements on different groups of migrants in the Netherlands, Germany and the UK, also concludes that the requirements do not have a positive effect on integration of any of the groups. Goodman and Wright (2015) examine the effect of language and knowledge requirements for immigration, settlement and citizenship on the political, social and economic integration of migrants, also find little evidence that these requirements produce tangible, long-term integration change (Goodman & Wright, 2015, p. 1885). Pochon-Berger and Lenz (2014), in a synthesis of the academic literature on language testing for immigration and integration purposes also summarize research findings saying that the requirement to pass such tests "is often judged as useless, stressful for the immigrants and discouraging, or it is criticized for selectively excluding certain groups (i.e. individuals with little formal education) from fuller integration" (Pochon-Berger & Lenz, 2014, p. 30). Indeed, there is no systematic research to date to support the assumption that language requirements have a positive effect on integration.

Test-theoretical reasons for language test misuse

So far, we have explored the linguistic and political reasons for test misuse. Another possible explanation is test theoretical in orientation.In the field of language testing and assessment, the concept of validity is essential; it defines what is considered the responsibility of professional language test developers and drives the research agenda. The question we explore here, is whether a marginalization of deliberate test misuse in validity theory and

---

[11]Comparison is feasible since most European countries relate their language demands to the CEFR levels.

validation frameworks may have caused language testers to ignore this practice, considering it outside of their professional responsibility. Indeed, for something to be considered a matter of relevance and concern to professional test developers and researchers, it needs to be encompassed in the very concept of validity (O'Sullivan & Weir, 2011). Herein lies a possible explanation for the lack of focus on deliberate test misuse in the field. We will explore this possible explanation by referring to some of the central theorists of the field and have a closer look at the theoretical debate about validity and test misuse in recent years.

Samuel Messick's conceptualisation of validity[12] presented in a series of publications in the 1980s and 90s (Messick, 1989, 1996, 1998) has been fundamental for the subsequent understanding and discussion of validity in educational measurement as well as in the professional language testing community. Messick's definition foregrounds the impact of tests on those affected by them as he states that validity: "[...] includes evidence and rationale for evaluating the intended and unintended consequences of score interpretation and use in both the short and long term [...]" (Messick, 1996, p. 251). Today, there is a fair degree of consensus that validity encompasses a concern for test consequences.[13] This is important, since: "[...] removing considerations of consequences from the domain of validity [...] would relegate them to a lower priority" (Linn, 1997, p. 16), a point echoed by Newton and Shaw (2016, p. 187). Importantly, despite his focus on test use and consequences, Messick explicitly argues that consequences which are *not* caused by deficiencies in the test instrument but by deliberate test misuse, fall outside the concept of validity and hence outside the scope of test validation. Defining deliberate test misuse outside of validity limits test developers' responsibilities, following Messick who argues that "test makers are not responsible for the consequences of misuse; the responsibility in this regard clearly lies with the (mis)user [...]". (Messick, 1998, p. 40). Influential validity frameworks building on Messick's definition also largely ignore deliberate test misuse.

Building on Messick's definition of Kane (1992), Kane (2013), 2016) and Bachman (2005) and Bachman, Palmer, and Palmer (2010) propose an argument-based approach for test validation, linking test scores and score-based interpretations to the uses made of test as well as to their consequences. The approach therefore appears ideal to validate language tests misuse. Nevertheless, there is a striking lack of actual validity studies applying the argument-based approach in the context of migration tests (Pochon-Berger & Lenz, 2014), and surprisingly enough, the immensely widespread use of language tests for migration regulation purposes with potentially negative consequences is hardly mentioned in Bachman & Palmer's most comprehensive work on validation (Bachman et al., 2010) and in recent publications on validity referring to it (Chapelle, 2020). We suggest three possible explanations for this apparent contradiction.

The underlying idea upon which an argument-based approach to validation rests, is that validity is not about inherent properties of a measurement instrument (such as the relation between test scores and the underlying construct of the test), but rather about the *strength of the arguments* (claims and warrants) put forward in support of the intended score interpretation and use (Chapelle, 2020). Test developers and test users "need to be able to demonstrate to stakeholders that the intended uses of their assessment are justified",

---

[12][...] an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment [...]. (Messick, 1989, p. 13).

[13]See Borsboom and Markus (2013) and Cizek (2020) for an opposing view.

Bachman et al. (2010, p. 2). Crucial to this approach is a scenario in which the test developer is very much in control of the whole process from test development to test use, interpretations and consequences. While this may be the case when scores are interpreted and used in line with the test developers' intentions, the argument-based approach seems less well-suited in the context of deliberate test misuse as described above. In these cases, the nature of the test construct as well as the uses and consequences of the test are "entirely externally determined as a function of policy and political processes" (McNamara, 2009, p. 161). In relation to detecting and preventing test misuse, it is an obvious weakness of the argument-based approach that validity depends on the strength of the arguments "for whatever test score interpretation and use one intends to defend" (Zumbo & Hubley, 2016, p. 300). Validity, then, becomes disconnected from the question of whether or not the test measures the construct. Rather, it is linked to whether or not test developers or test users manage to build a convincing argument for their use (Chalhoub-Deville & O'Sullivan, 2020).

Another central aspect of argument-based validation is the degree to which the actual uses and consequences of the test matches the *intended uses and consequences*. As argued earlier in this paper, there is reason to believe that the purpose of language tests for migration purposes is exclusion of certain groups of migrants, regardless of whether or not policy makers state this intention explicitly. When language requirements were introduced and in Norway and the level of proficiency later raised, the explicitly stated aim was to reduce the number of migrants and to ensure that citizenship is something migrants have to make a real effort to achieve (Kunnskapsdepartementet, 2020). Indeed, after the introduction of the test for citizenship, the number of new citizens dropped to under half of the number of the preceding year.[14] The effect of the test matched the intention. Does this then make the test valid? Following the logic of the argument-based approach, it is hard to see how detrimental consequences in and of themselves invalidate a test or test use, as long as the actual consequences match the intended consequences. Several researchers have pointed to the problems of linking validity to the intended interpretations, uses and consequences and have called for a focus on how tests and test scores are *actually* interpreted and used and what their consequences are in practice (Cronbach, 1988; Slomp, Corrigan, & Sugimoto, 2014; Moss, 2016; Newton & Shaw, 2016, p. 181; Shohamy, 2009). Important to note, this challenge is not solved by claiming that one should investigate both the intended and the *unintended* consequences: As we saw above, policy makers are often not explicit about their intended consequences, their agendas may very well be hidden (Shohamy, 2006). The solution that we propose lies in a wider definition of test misuse to include negative consequences regardless of whether they are intentional or unintentional, and that misuse be included in the definition of validity. We will return to this point in the last part of the paper.

This brings us to our final point, the *question of beneficence*. Central in the argument-based approach is the claim that: "[t]he *consequences* of using an assessment and of the decisions that are made are beneficial to all stakeholders (Bachman et al., 2010, p. 103)". Indeed, Bachman & Palmer begin with "[t]he premise that people generally use language assessment in order to bring about some beneficial outcomes or consequences for stakeholders [...]" (ibid, p. 86). This deeply optimistic view of language test users' good intentions is however in stark contrast with the use of language tests to regulate migration

---

[14]Similar effects are found in other countries after the introduction of such tests (Oers, 2020).

in the Western world today (Slade, 2010; McNamara & Roever, 2006, p. 38). The crucial questions in relation to beneficence are; *What is beneficial? Beneficial to whom?*, and *Beneficial according to whom?*, but these important questions are not further debated in Bachman et al. (2010). As McNamara convincingly argues, the answers to these questions depend on who you ask (McNamara, 2005, p. 356). We cannot assume that the intended purpose of test and test use is to bring about beneficial outcomes for all stakeholders, which may not even be possible. Rather, it is likely to assume that for those failing the test the consequences will be less beneficial, or even harmful. In addition, one must assume that policy makers introducing language tests to control migration do so in order to achieve what they would argue to be beneficial. Many countries introduced migration tests as part of a stricter immigration legislation following 9/11 and the subsequent bombings in London and Madrid, hence the purpose of introducing requirements could be argued to be beneficial in that they contribute to securing social cohesion, reducing the fear and risk of terror (Slade, 2010). For beneficence to be more than a relative term for all including the most vulnerable test takers, it needs to be closely linked to test ethics and fundamental human and democratic rights of those subject to them.

Can test misuse be prevented?

In this paper we have pointed to a use of language tests which has become widespread, and we have shown that language tests are used purposefully to regulate migration and limit access to entry, residency and citizenship, a practice which is found to be non-beneficial to those migrants who are unable to meet the requirements. We have shown that when strict requirements are introduced, low literate adult migrants are particularly challenged. We have defined test misuse as a use of tests with non-beneficial or harmful consequences for test takers, regardless of whether these consequences are intentional or not. Finally, we have looked at linguistic, political and test-theoretical explanation why language tests are being subject to political hijacking. What remains to be said, is whether there is a way forward in order to prevent deliberate test misuse of the kinds discussed here, and in particular, what professional language testers can do.

First and foremost – for language testers to engage actively in trying to prevent language tests from being misused, they need to see it as their professional responsibility: If test misuse is defined only with respect to whether tests are used beyond their intentional purposes, deliberate political use of language tests for gatekeeping purposes falls outside the definition. We therefore strongly support Shohamy and others' view of test misuse referring to the use of a test yielding harmful consequences for test takers, regardless of whether or not these consequences were intentional. The first step towards a stronger conception of ethical responsibility from test developers is a redefinition of test misuse as presented in this paper.Secondly, for test misuse to be considered relevant for language testers and test researchers, it needs to be incorporated in validity theory and validation frameworks. The concept of validity and validation frameworks building on it, need to be broadened in two respects: Test consequences of relevance to validity should not be restricted to consequences caused by flaws in the construct, as argued by Messick (1989, 1998). Consequences need to be put to the front of validation studies, regardless of whether or not the test as such holds high empirical qualities. Moreover, validation frameworks need to be broadened to incorporate a focus on negative consequences caused by the test, no

matter what the original intention of test developers or policy makers were, hence, to include test misuse in all forms.

As discussed above, the argument-based approach rests upon an optimistic assumption that tests are developed and used with the purpose of bringing about beneficial consequences for stakeholders. This positive assumption is a poor reflection of the way language tests are used for gatekeeping and control in migration contexts: We need to be realistic and consider the possibility that the interpretations and use of our tests by policy makers may not be in the best interest of test-takers. For individuals less likely to meet the requirements, the impact may be detrimental.

The professional field of language testing and assessment needs a validation framework that is suitable not only for language testers to justify the intended uses and consequences of their tests, but to demonstrate and speak up against harmful consequences, also when these are the effect of a deliberate policy of exclusion and control, which as we saw above, may be explicitly stated, or hidden behind an alternative agenda of promoting integration. A validity framework, which in our opinion, has the potential to do just that, is the socio-cognitive (SC) model proposed by O'Sullivan and Weir (2011) and O'Sullivan (2016) and presented in detail in the recent book on validity by Chalhoub-Deville and O'Sullivan (2020). The SC-model places the test takers with their physical, psychological and experiential characteristics in the centre of attention in validation studies, thereby catering for the concern for test-taker group differences, like the ones described in relation to LESLLA learners above. In addition, the SC-model emphasizes test consequences and argues that validity always needs to be investigated in close connection to the local context and use, referred to as localisation.

If language test developers and researchers see it as their professional responsibility to work to prevent test misuse as described above, we are optimistic as to what can be achieved. Even though test developers cannot decide how their tests are to be used, they can engage more actively in language test advocacy on different areas related to test development and test research, for example in taking an active part in public debate speaking out against test misuse and injustice, informing public opinion about the meaning of test scores and proficiency levels as well as about more or less appropriate use of tests, engaging in dialogue with policy makers, stressing the importance of needs analyses, as well as conducting research into the consequences of certain uses of language tests for different groups of learners, especially LESLLA learners. With this paper, we wanted to raise awareness of test misuse and highlight the danger of marginalizing deliberate test misuse in validation theory, since a theoretical marginalisation may imply a *de facto* abandonment of the issue in language testing research and practice. We hope this paper can serve the purpose of putting language test misuse on the agenda and raise language test developers' awareness about their responsibility in preventing misuse.

Disclosure statement

ORCID

Cecilie Hamnes Carlsen ⓘ http://orcid.org/0000-0001-5130-8206
Lorenzo Rocca ⓘ http://orcid.org/0000-0002-0504-4868

References

Allemano, J. (2013). Testing the reading ability of low-educated ESOL learners. *Apples-Journal of Applied Language Studies*, *7*(1), 67–81.

ALTE. (2020). *ALTE principles of good practice*. Cambridge, UK: ALTE Secretariat.

Anderson, B. (2016). *Imagined communities* (3rd ed.). London, UK: Verso.

Andringa, S., & Godfroid, A. (2019, January 25). *SLA for all? Reproducing SLA research in non-academic samples*. osf.io/mp47b. https://doi.org/10.17605/OSF.IO/MP47B

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, UK: Oxford University Press.

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly: An International Journal*, *2*(1), 1–34. doi:10.1207/s15434311laq0201_1

Bachman, L. F., Palmer, A. S., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford, UK: Oxford University Press.

Blackledge, A. (2009). "As a country we do expect": The further extension of language testing regimes in the United Kingdom. *Language Assessment Quarterly*, *6*(1), 6–16. doi:10.1080/15434300802606465

Borsboom, D., & Markus, K. A. (2013). Truth and evidence in validity theory. *Journal of Educational Measurement*, *50*(1), 110–114. doi:10.1111/jedm.12006

Bruzos, A., Erdocia, I., & Khan, K. (2018). The path to naturalization in Spain: Old ideologies, new language testing regimes and the problem of test use. *Language Policy*, *17*(4), 419–441. doi:10.1007/s10993-017-9452-4

Canale, M. & Swain, M. (1980). Theoretical bases of communiative approaches to second language teaching and testing. *Applied Linguistics 1*, 1–47.

Carlsen, C. H. (2017). Giving LESLLA-learners a fair chance in testing. In M. Sosinski (Ed.), *Proceedings from LESLLA 2016 12th annual symposium, 8–10 September 2016* (pp. 135–148), Granada, Spain: Universidad de Granada.

Celce-Murcia, M. (2008). Rethinking the role of communicative competence in language teaching. In E. A. Soler & M. S. Jordà (Eds.), *Intercultural language use and language learning*. Heidelberg, Germany: Springer.

Chalhoub-Deville, M., & O'Sullivan, B. (2020). *Validity: Theoretical development and integrated argument*. Sheffield, UK: British Council Monographs.

Chapelle, C. (2020). *Argument-based validation in testing and assessment*. Los Angeles, CA: Sage.

Cizek, G. J. (2020). *Validity: An integrated approach to test score meaning and use*. Abingdon Oxon, UK: Routledge.

Council of Europe. (2001). *The common European framework of reference for languages*. Cambridge, UK: Cambridge University Press.

Council of Europe. (2014). Parliamentary assembly of the council of europe recommendation 2034

Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment – Companion volume with new descriptors*. Cambridge, UK: Author.

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 3–17). Hillsdale, NJ: Lawrence Erlbaum Associates.

Davis, A. (2012). Ethical codes and unexpected consequences. In G. Fulcher & F. Davidson (Eds.), *The routledge handbook of language testing* (pp. 455–468). New York, NY: Routledge.

Deygers, B., Zeidler, B., Vilcu, D., & Carlsen, C. H. (2018). One framework to unite them all? Use of the CEFR in European university entrance policies. *Language Assessment Quarterly*, *15*(1), 3–15. doi:10.1080/15434303.2016.1261350

Edwards, J. (2009). *Language and identity: An introduction*. Cambridge: Cambridge University Press.

Evans, D. (Ed.). (2019). *Language, Identity and symbolic culture*. London, UK: Bloomsbury

Extra, G., & Spotti, M. (2009). Testing regimes for newcomers to the Netherlands. In G. Extra, M. Spotti, & P. Van Avermaet (Eds.), *Language Testing, Migration and Citizenship. Cross-national Perspectives on Integration Regimes* (pp. 128–147). London, UK: Advances in Sociolinguistics.

Fulcher, G. (2000). The «communicative» legacy in language testing. *System*, *28*(4), 483–497. doi:10.1016/S0346-251X(00)00033-6

Fulcher, G., & Davidson, F. (2009). Test architecture, test retrofit. *Language Testing*, *29*(1), 123–144. doi:10.1177/0265532208097339

Goodman, S. W. (2017). *Immigration and membership politics in Western Europe*. Cambridge, UK: Cambridge University Press.

Goodman, S. W., & Wright, M. (2015). Does mandatory integration matter? Effects of civic require-ments on immigrant socio-economic and political outcomes. *Journal of Ethnic and Migration Studies*, *41*(12), 1885–1908. doi:10.1080/1369183X.2015.1042434

Hulstijn, J. (2015). *Language proficiency in native and non-native speakers – Theory and research*. Amsterdam, The Netherlands: John Benjamins Publication Company.

Hymes, D. H. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269–293). Harmondsworth, UK: Penguin.

Joppke, C. (2010). *Citizenship and Immigration*. Cambridge, UK: Polity Press.

Kane, M. (1992). An argument-based approach to validity. *Psychological Bulletin*, *112*(3), 527. doi:10.1037/0033-2909.112.3.527

Kane, M. (2016). Explicating validity. Assessment in education: Principles. *Policy & Practice*, *23*(2), 198–211.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1–73. doi:10.1111/jedm.12000

Kim, J. W., Yoon, J., Kim, S. R., & Kim, H. H. (2014). Effect of literacy level on cognitive and language tests in Korean illiterate older adults. *Geriatrics & Gerontology International*, *14*(4), 911–917. doi:10.1111/ggi.12195

Kunnan, A. (2009). Testing for Citizenship: The U.S. Naturalization Test. *Language Assessment Quarterly*, *6*(1), 89–97. doi:10.1080/15434300802606630

Kunnskapsdepartementet. (2020). Proposisjon 98L. (2019–2020). Endringer i statsborgerloven (krav om ferdigheter i norsk muntlig) ["Changes in the citizenship law – Requirements of proficiency in oral Norwegian"].

Kurvers, J., Van De Craats, I., & van Hout, R. (2015). Footprints for the future: Cognition, literacy and second language learning by adults. In I. Van De Craats, J. Kurvers, & R. van Hout (Eds.), *Adult literacy, second language, and cognition*, (pp. 7–32). Nijmegen, The Netherlands..

Linn, R. L. (1997). Evaluating the validity of assessments: The consequences of use. *Educational Measurement: Issues and Practice*, *16*(2), 14–16. doi:10.1111/j.1745-3992.1997.tb00587.x

Mackenzie, C. (2010). Citizenship, identity, and immigration: contemporary philosophical perspec-tives. In C. Slade & M. Möllering (Eds.), *From migrant to citizen* (pp. 191–216). London, UK: Palgrave Macmillan

McNamara, T. (2005). 21st Century Shibboleth: Language tests, identity and intergroup conflict. *Language Policy*, *4*(4), 351–370. doi:10.1007/s10993-005-2886-0

McNamara, T. (2009). Language tests and social policy. A commentary. In Hogan-Brun, G., Mar-Molinero, C. & Stevenson, P. (Eds.). Discourses on Language and Integration. *Critical perspectives on language testing regimes in Europe*. Amsterdam, The Netherlands: John Benjamins Publishing Company

McNamara, T. (2010). The use of language tests in the service of policy: Issues of validity. *Revue française de linguistique appliquée*, *1*(XV), 7–23. doi:10.3917/rfla.151.0007

McNamara, T. (2020). The Anti-Shibboleth: The traumatic character of the shibboleth as silence. *Applied Linguistics*, *41*(3), 334–351. doi:10.1093/applin/amaa007

McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA: Blackwell Publishing.

McNamara, T., & Shohamy, E. (2008). Language tests and human rights. *International Journal of Applied Linguistics*, *18*(1), 89–95. doi:10.1111/j.1473-4192.2008.00191.x

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed. pp. 13–104). New York, NY: American Council on education and Mcmillan.

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, *13*(3), 241–256. doi:10.1177/026553229601300302

Messick, S. (1998). Validity: A matter of consequences. *Social Indicators Research*, *45*(1–3), 35–44. doi:10.1023/A:1006964925094

Midtbøen, A. (2015). Citizenship, integration and the quest for social cohesion: Nationality refor in the Scandinavian countries. *Comparative Migration Studies*, *3*(3), 1–15. doi:10.1007/s40878-015-0002-y

Moss, P. A. (2016). Shifting the focus of validity for test use. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 236–251.

Newton, P. E., & Shaw, S. D. (2016). Disagreement over the best way to use the word 'validity' and options for reaching consensus. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 178–197.

Norton, B. (2013). *Identity and language learning*. Bristol, UK: Multilingual Matters.

O'Sullivan, B., & Weir, C. (2011). Language testing and validation. In B. O'Sullivan (Ed.), *Language testing theories and practices* (pp. 13–32). New York, USA: Palgrave Macmillan.

O'Sullivan, B. (2016). Adapting tests to the local context. In *New directions in language assessment, special edition of the JASELE Journal* (pp. 145–158). Tokyo, Japan: Japan Society of English Language Education & the British Council.

Oers, R. van (2014). Deserving Citizenship. Citizenship Tests in Germany, the Netherlands and the UK. M. Nijhoff

Oers, R., Ersbûll, E., & Kostakopoulou, T.. (2010) Mapping the redefinition of belonging in Europe. In R. Oers, E. Ersbûll, T. Kostakopoulou, & N. Radboud Universiteit (Ed.), *Centre for Migration Law. A re-definition of belonging?: Language and integration tests in Europe* (Vol. 20, pp. 307–332). M. Nijhoff. Leiden, The Netherlands: Nijhoff eBook titles.

Oers, R.. (2020) Deserving citzenship in Germany and The Netherlands. Citizenship tests in liberal democracies. *Etnicities. Special Issue: Naturalization Policies*, *21*(2), 1–18.https://doi.org/10.1177%2F1468796820965785

Pochon-Berger, E., & Lenz, P. (2014). *Language requirements and language testing for immigration purposes. A synthesis of academic literature*. Report of the Research Freiburg, Germany: Centre on Multilingualism.

regjeringen. (2015) Tiltak for å møte flyktningkrisen ("Measures to meet the refugee crisis"). https://www.regjeringen.no/no/aktuelt/tiltak-for-a-mote-flyktningkrisen/id2469066/

Rocca, L., Deygers, B., & Carlsen, C. H. (2020). *Linguistic integration of adult migrants: Requirements and learning opportunities*. Strasbourg, France: Council of Europe publication.

Shohamy, E. (1998). Critical language testing and beyond. *Studies in Educational Evaluation*, *24*(4), 331–345. doi:10.1016/S0191-491X(98)00020-0

Shohamy, E. (2001). *The power of tests. A Critical Perspective on the Use of Language Tests*. Harlow, UK: Pearson.

Shohamy, E. (2006). *Language Policy: Hidden Agendas and New Approaches*. Oxon, UK: Routledge.

Shohamy, E. (2009). Language tests for immigrants. Why language? Why tests? Why citizenship? In Hogan-Brun, C. Mar-Molinero, & P. Stevenson (Eds.), *Discourses on Language and Integration: Critical perspectives on language testing regimes in Europe*(pp. 45–60). Amsterdam, UK: John Benjamins Publishing Company.

Shohamy, E., & Menken, K. (2015). Language Assessment. Past to present misuses and future possibilities. In W. Wright, S. Boun, & O. Garcia (Eds.), *The Handbook of Bilingual and Multilingual Education* (pp. 253–269). Oxford, UK: Wiley Blackwell.

Slade, C. (2010). Shifting Landscapes of Citizenship. In C. Slade & M. Möllering (Eds.), *From Migrant to Citizen: Testing Language, Testing Culture* (pp. 3–23). Hampshire, UK: Palgrave Macmillan.

Slade, C., & Möllering, M.2010 (Eds.). *From migrant to Citizen: Testing language, testing culture*. Hampshire, UK: Palgrave Macmillan.

Slomp, D., Corrigan, J., & Sugimoto, T. (2014). A framework or using consequential validity evidence in evaluating large-scale writing assessment: A Canadian study. *Research in the Teaching of English*, *48*(3), 276–302.

Strik, T., Böcker, A., Luiten, M., & Oers, R. van (2010). *The INTEC Project: Integration and naturalisation tests: The new way to European citizenship*. Nijmegen, The Netherlands: Radboud University.

Tarone, E. (2010). Second language acquisition by low-literate learners: An under-studied population. *Language Teaching*, *43*(1), 75–83. doi:10.1017/S0261444809005734

UNESCO. (2017). *Literacy rates continue to rise from one generation to the next* FS/2017/LIT/45).

Van Avermaet, P., & Gysen, S. (2009). One nation two policies: Language requirements for citizenship and integration in Belgium. In G. Extra, M. Spotti, & P. Van Avermaet (Eds.), *Language testing, migration and citizenship. Cross-national perspectives on integration regimes* (pp. 107–124). London, UK: Advances in Sociolinguistics.

Van Avermaet, P., & Pulinx, R. (2014). Language testing for immigration to Europe. In A. Kunna (Ed.), *The companion to language assessment* (pp. 1–14). Malden, USA: John Wiley & Sons, Inc..

Van De Craats, I., Kurvers, J., & Young-Scholten, M. (2006). Research on low-educated second language and literacy acquisition. *LOT Occasional Series*, *6*, 7–23.

Wall, D. (2012). Washback. In G. Fulcher & F. Davidson (Eds.), *The Routledge Handbook of Language Testing* (pp. 79–92). Oxon, UK: Routledge.

Yosso, T. J. (2005). Whose culture has capital? A critical race theory discussion of community cultural wealth. *Race Ethnicity and Education*, *8*(1), 69–91. doi:10.1080/1361332052000341006

Zumbo, B., & Hubley, A. (2016). Bringing consequences and side effects of testing and assessment to the foreground. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 299–303.