Why bother with testing? The validity of immigrants' self-assessed language proficiency

Aileen Edele [a,*], Julian Seuring [b], Cornelia Kristen [b], Petra Stanat [c]

[a] *Humboldt-Universität zu Berlin, Germany*
[b] *Chair for Sociology, esp. Analysis of Social Structure, University of Bamberg, Germany*
[c] *Institute for Educational Quality Improvement, Humboldt-Universität zu Berlin, Germany*

A R T I C L E   I N F O

A B S T R A C T

Due to its central role in social integration, immigrants' language proficiency is a matter of considerable societal concern and scientific interest. This study examines whether commonly applied self-assessments of linguistic skills yield results that are similar to those of competence tests and thus whether these self-assessments are valid measures of language proficiency. Analyses of data for immigrant youth reveal moderate correlations between language test scores and two types of self-assessments (general ability estimates and concrete performance estimates) for the participants' first and second languages. More importantly, multiple regression models using self-assessments and models using test scores yield different results. This finding holds true for a variety of analyses and for both types of self-assessments. Our findings further suggest that self-assessed language skills are systematically biased in certain groups. Subjective measures thus seem to be inadequate estimates of language skills, and future research should use them with caution when research questions pertain to actual language skills rather than self-perceptions.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Across disciplines, research on immigrants and their descendants[1] views language skills as a key to their educational, economic and social integration into receiving countries (Chiswick and Miller, 2001; Jimenez, 2011; Portes and Rumbaut, 2006). A substantial body of research has examined the conditions for acquiring the language of the country of residence (second language, L2) and for the maintenance or loss of the language spoken in the country of origin (first language, L1) (e.g., Alba et al., 2002; Carliner, 2000; Chiswick and Miller, 2001; Van Tubergen and Kalmijn, 2009). In addition, various studies have aimed to determine the effects of immigrants' language skills in both L1 and L2 on core aspects of their integration, such as educational attainment (e.g., Bankston and Zhou, 1995; Lutz and Crist, 2009; Mouw and Xie, 1999; Nielsen and Lerner, 1986) or economic success (e.g., Chiswick and Miller, 1995; Dustmann and Fabbri, 2003; Hwang et al., 2010; Kaida, 2013).

Whereas studies of smaller samples sometimes apply psychometric tests to assess immigrants' linguistic skills, population and household censuses as well as large-scale studies typically measure participants' language proficiency with self-assessments. A substantial proportion of research on immigrants' language proficiency, particularly in sociology and

---

* Corresponding author at: Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany.
E-mail address: aileen.edele@iqb.hu-berlin.de (A. Edele).
[1] The term "immigrant," as it is used in this paper, refers to immigrants (first generation) and their children (second and third generations).

economics but also in other fields, relies on data from these studies and thus on self-reports (e.g., Berry et al., 2006; Carliner, 2000; Chiswick and Miller, 1995; Chiswick et al., 2004; Mouw and Xie, 1999; Pendakur and Pendakur, 2002; Van Tubergen and Kalmijn, 2009).

Self-assessments (or self-estimates) of ability (or skill) reflect a person's self-concept, which can be broadly defined as "a person's perception of himself" (Shavelson et al., 1976, p. 411) in a particular domain. Whereas earlier views of this construct focused on the global self-concept, domain-specific conceptualizations now prevail (e.g., Bong and Skaalvik, 2003; Marsh, 1990, 1992). As psychological research shows, self-concepts of ability correlate substantially with actual levels of ability in a given domain (Marsh, 1992). Whether self-assessed skills present valid indicators of actual skills, however, is questionable, and the extent to which self-assessments of language skills measure the same (or at least very similar) skills as language tests is unclear. In other words, we cannot presume that, for example, investigations using self-assessed language skills as predictors of immigrants' educational outcomes will arrive at the same conclusions as analyses using language tests.

The main purpose of the present study is to determine whether language self-assessments of the type that is frequently used in research on immigrants yield results similar to those of competence tests. Our analyses draw on a large sample of immigrant youth participating in the German National Educational Panel Study (NEPS), for which both self-assessed language skills and tested language abilities in L1 and L2 are available. The first set of analyses focuses on L2 abilities and includes all students of immigrant descent assessed in the 9th-grade sample of the NEPS (4078 students in total). The families of the two largest groups of immigrant students immigrated to Germany from the Former Soviet Union or from Turkey, and the number of students of Polish descent is also substantial. Reflecting the heterogeneity of the immigrant population in Germany, however, most students are of other origins. The second set of analyses focuses on immigrant students' L1 ability and is based on a subsample of Russian- or Turkish-speaking students (1164 students in total).

In a first step, we correlate language self-assessments with performance on language tests, assuming that the language tests provide valid estimates of participants' actual linguistic skills. We then move beyond this bivariate approach and use self-assessed language skills vs. language test scores in a number of regression models pertaining to substantive research questions to compare the findings. The validity of self-assessments is supported if these models yield similar findings, whereas substantial differences in the results indicate that self-assessments are of limited validity.

Determining the validity of self-assessed language skills is important for at least two reasons. First, it provides an indication of the extent to which the findings from the large number of previous studies that have relied on self-assessments are valid (e.g., Berry et al., 2006; Carliner, 2000; Chiswick and Miller, 1995; Mouw and Xie, 1999; Pendakur and Pendakur, 2002; Van Tubergen and Kalmijn, 2009). Second, our study provides important guidance for designing future research. Because testing is time consuming and expensive, researchers typically prefer to use self-assessments. However, this route should only be taken if the measures applied are valid.

2. Evidence for the validity of ability self-assessments

In general, the accuracy of ability self-assessments is questionable. It is not just an amusing *bon mot* but an empirical finding that eight out of ten drivers report that they have above-average driving skills (McCormick et al., 1986). Driving is certainly not the only domain in which people have difficulty adequately estimating their own abilities. Meta-analyses examining the correlation between ability self-assessments and test scores or teacher ratings in various domains indicate that self-assessments are often inaccurate (Falchikov and Boud, 1989; Freund and Kasten, 2012; Hansford and Hattie, 1982; Mabe and West, 1982). These studies find mean correlations that are positive, yet moderate (typically ranging between .3 and .4) with considerable variability in the correlation coefficients across studies. Although some researchers argue that ability self-assessments are accurate (e.g., Ackerman et al., 2002), most scientists agree that their validity is limited (e.g., Carter and Dunning, 2008; Dunning et al., 2004; Kruger and Dunning, 1999).

The self-assessment measures that have been implemented in empirical studies vary substantially. An important distinguishing feature is their degree of specificity. Some self-assessments refer to broad domains and underlying abilities that affect behavior across situations and over time, such as scientific or reading ability, and are thus quite general. Other self-assessments focus on more specific skills, such as understanding a science article, or refer to performance on a concrete task that participants just completed in the same study or typically encounter in everyday situations. Many researchers assume that self-assessment measures with high degrees of specificity are more likely to evoke accurate responses because they are less ambiguous than more general measures (Ackerman et al., 2002; Freund and Kasten, 2012; Mabe and West, 1982; Marsh, 1990). Empirical findings support this assumption. Kruger and Dunning (1999), for instance, found a substantial correlation between performance on a grammar test and a self-estimated test score, whereas participants' self-assessments of their general grammar ability were not significantly related to their test performance. Integrating the findings from a larger number of studies, a meta-analysis confirmed that self-assessed performance on specific tasks relates more strongly to the validation criteria than to self-assessed underlying abilities (Mabe and West, 1982). Thus,