Who succeeds and who fails? Exploring the role of background variables in explaining the outcomes of L2 language tests

Ann-Kristin Helland Gujord (iD)

University of Bergen, Norway

Abstract

This study explores whether and to what extent the background information supplied by 10,155 immigrants who took an official language test in Norwegian affected their chances of passing one, two, or all three parts of the test. The background information included in the analysis was prior education, region (location of their home country), language (first language [L1] background, knowledge of English), second language (hours of second language [L2] instruction, L2 use), L1 community (years of residence, contact with L1 speakers), age, and gender. An ordered logistic regression analysis revealed that eight of the hypothesised explanatory variables significantly impacted the dependent variable (test result). Several of the significant variables relate to pre-immigration conditions, such as educational opportunities earlier in life. The findings have implications for language testing and also, to some extent, understanding of variation in learning outcomes.

Introduction

Adult migrants learning a second language (L2) are a heterogeneous group of people who (for a variety of reasons) have left their home country and found themselves in a situation where they need to learn the officially recognised language of communication in a new society. Some are economic immigrants (skilled workers with varying levels of training and education), others have migrated because of education, some are refugees

Corresponding author:
Ann-Kristin Helland Gujord, Department of Linguistic, Literary and Aesthetic Studies, University of Bergen, Postboks 7805, 5020 Bergen, Norway.
Email: ann-kristin.gujord@uib.no

escaping persecution or inhumane treatment in their home country or who have been granted international protection on other humanitarian and compassionate grounds, and others have left home due to family migration. These groups of people from diverse backgrounds bring a large variation in social, cultural, and educational context to the language testing situation.

In this study, I explore whether the background information supplied by adult immigrants who took an official language test in Norwegian affects their test outcome. Background variables have, to a certain extent, been the focus of language testing research, as the impact of *test taker characteristics* (Bachman, 1990; Kunnan, 1995) or *personal characteristics* (Bachman & Palmer, 1996) includes variables such as educational background (Davies et al., 1999). Understanding variation in learning outcomes is one of the core issues in a field closely related to the field of language testing—second language acquisition (SLA)—as a major distinguishing characteristic of adults' L2 learning is the varying degree of success. Due to the recent introduction of language requirements for citizenship and permanent residency in many countries worldwide, the implications of *not* learning the dominant language sufficiently to meet the language requirements are becoming even more severe (Extra et al., 2009; Hogan-Brun et al., 2009; Rocca et al., 2020; Shohamy & McNamara, 2009). Accordingly, more research on the role of background variables, or test taker/personal characteristics, in understanding variation in test outcomes, is important. This is not relevant only for the field of language testing, but also for SLA research, as exploration of the large data sets available in testing programmes can potentially expand the knowledge base on sources of individual variation in L2 learning.

In this study, the purpose is to gain new insights into the explanatory variables of individual variation in test outcomes using a large-scale data set of 10,155 adult learners of Norwegian as the L2, who have taken an official test in the Norwegian language for immigrants. The test measures basic language skills (level A2), defined according to the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001). On the day of the test, candidates completed a form that requested individual background information for each test taker, and some of this information comprised the hypothesised explanatory variables (independent variables) in this study. The analysis was conducted using an ordered logistic regression that aimed to explore the extent to which background variables explain variation in test result.

Research on the role of background variables in language testing and SLA research

In language testing research, there has been an interest in variables that influence test performance (Bachman, 1990; Kunnan, 1995), and in distinguishing factors that are desirable, that is, relevant to the concept measured, from factors that are not desirable, and which therefore represent sources of measurement error. Some of this research is related to the process of validating language tests to isolate and calculate construct-irrelevant variance, which occurs "when the test contains excess reliable variance that is irrelevant to the interpreted construct" (Messick, 1989, p. 34). Some key sources of construct-irrelevant variance in an L2 proficiency test are *test taker characteristics*

(Bachman, 1990) or *personal characteristics* (Bachman & Palmer, 1996). Bachman and Palmer (1996) define personal characteristics as "individual attributes that are not part of test takers' language ability, but which may still influence their performance on language tests" (p. 64). Many types of *test taker characteristics* examined in the literature in the 1980s and 1990s, when this concept was rather heavily researched, were learner related —such as language aptitude, personality, cognitive style (Bachman & Palmer, 1996), attitude or motivation (Kunnan, 1995)—or were demographic variables such as age, sex (Davies et al., 1999) and "affective reactions to test taking" (Davies et al., 1999, p. 208). Yet, variables related to test takers' backgrounds, such as cultural background, language background, educational background and background knowledge, were also regarded as attributes irrelevant to language tests (Davies et al., 1999).

Test taker/personal characteristics remain understudied, despite their relevance for the validity of language tests. As noted by O'Sullivan and Green (2011, p. 37), "Current concerns with test fairness, which emphasise the social responsibility of language testers, have highlighted the need to avoid bias against certain test takers." The concept of fairness is undoubtedly relevant, as it is an essential quality of language tests that ensures equal treatment of test takers. Fairness (and justice) in language testing is a topic that has received attention in language testing research since the turn of the century. This increase in focus is closely connected to the introduction of obligatory language and knowledge of society tests for migrants applying for residency and citizenship worldwide, which has led many language testing researchers to critically evaluate the consequences of such language requirements as well as the rationale behind migration tests (Extra et al., 2009; Hogan-Brun et al., 2009; Rocca et al., 2020; Shohamy & McNamara, 2009). Studies of background variables that influence test performance could yield important insights for identifying learner groups that, because of their previous experience or lack thereof (e.g. with schooling), have specific challenges that may impact their performance. Accordingly, language testers must be aware of these challenges when constructing tests that involve language abilities (see, for example, Carlsen, 2017).

Variation in *test outcome* can be understood as an indication of variation in *learning outcome*, as a language test result is the representation of an assessment of a language proficiency level. While some L2 learners achieve advanced levels of proficiency in the target language, other learners' language abilities stabilise at a state characterised by simple grammatical structures, limited vocabulary, and systematic errors. However, in SLA research, little attention has been given to the role of background variables, that is, factors that have largely been determined before the immigrants entered the country and started learning the L2. The SLA subfield that has most systematically investigated the sources of individual variation, individual differences research, has primarily studied learner-*internal* features, "intrinsic abilities and personality dimension[s]" (Dewaele, 2009, p. 24), for instance personality, language aptitude, motivation, learning style, and learning strategies (Dörnyei, 2005).

The role of learner-*external* factors in understanding individual variation is emphasised in another vein of SLA: research that conceives language learning as a fundamentally social phenomenon that must be examined in the light of the learning context on both micro- and macro-levels. For instance, Norton and Toohey (2001) explicitly discussed how social-oriented approaches to SLA can shed light on variation in learning

outcomes. They conclude that to understand individual variation, emphasis must be placed not only on what the individual L2 learners do but also on "social practices in the contexts in which individuals learn L2s" (p. 318). Additional L2 studies also connect individual variation to macrostructures. Darvin and Norton (2014) suggested that differences in socio-economic class provided varying opportunities for learning and thereby diverse learning outcomes for two Filipinos from different class backgrounds studying in Canada. Although this type of social-oriented SLA research shows that variation in learning outcomes must also be understood outside a learner-internal perspective, the focus of this research is still on the environment in the L2 community and the conditions for L2 learning after migration. Learners' social backgrounds before migration are barely given attention.

However, over the last decade, a focus on less educated migrants has slightly strengthened the focus on the importance of L2 learners' backgrounds to understand variation in learning outcomes. This research is closely connected to the Literacy Education and Second Language Learning for Adults forum (LESLLA),[1] which was founded in 2005 and aims to promote research findings on how low or non-literate adults with, at the most, primary schooling in their L1s, learn languages (as well as pedagogical and political implications of such research). An important impetus for research on low-literate/low-educated L2 learners is the sampling bias that has recently been uncovered and discussed in the SLA research community, namely an overreliance on learners from western, educated, industrialised, rich, and democratic (WEIRD) countries (Henrich et al., 2010) and a lack of study samples comprising illiterate or non-academic L2 learners (e.g. Ortega, 2005; Tarone & Bigelow, 2005; Van de Craats et al., 2006; Young-Schoulten, 2013). This is unfortunate because the average educational level of immigrants to the US and European states tends to be notably lower than that of the host country's population (European Commission-Organisation for Economic Co-operation and Development [EU-OECD], 2016). European statistics also reveal that host country language command is positively correlated with educational background, and less educated immigrants (most often refugees) represent the most vulnerable groups of immigrants in the labour market (EU-OECD, 2016; Organisation for Economic Cooperation and Development [OECD], 2019). Yet, there are few L2 studies that explicitly explore the impact of educational background on L2 learning.

There are, however, a few exceptions. In a recent study on differences in the learning progress of 24 adult learners (over 40 years old) of L2 English in Australia, their prior level of education was found to be associated with "considerable language learning gains" (Kozar & Yates, 2019, p. 181). Strube et al. (2013) studied the learning processes in L2 literacy classes in the Netherlands and referred to three major projects that focused on L2 learning in low-literate/low-educated adult learners: *What Works* in the USA (Condelli et al., 2003), *ESOL effective teaching and learning* in Great Britain (Baynham et al., 2007), and *Alfabetisering NT2 in beeld: Leerlast en succesfactoren* [Focus on L2 literacy: Study load and success factors] in the Netherlands (Kurvers & Stockmann, 2009). In some of these studies, the influence of background variables (learner characteristics) on the learning process and learning outcomes was analysed. According to Strube et al.'s (2013) summary, these studies showed that prior education is a success factor for L2 learning among this specific learner group, together with age and degree of contact

with L1 speakers (p. 47). However, two of the studies found different effects on formal instruction. Kurvers and Stockmann (2009) found a negative correlation between classroom hours and reading skills, whereas Baynham et al. (2007) found a moderate positive correlation. Importantly, in all of these studies, the rate of attendance was significant for language performance (Strube et al., 2013, p. 62). In their longitudinal study, Strube et al. found that age was an important variable for understanding differences among six learners' oral development in the L2 (p. 61). Neither hours of classroom instruction nor the rate of attendance significantly correlated with learning progress measures (i.e. pre- and post-assessment design; principal component analysis, Strube et al., 2013, p. 52).

Outside SLA research and within the wider context of research on immigration and integration, there are large-scale studies that provide insight into the impact of various factors related to immigrant background, for instance, educational background, age, and degree of contact with L1 speakers. For instance, in a study of male immigrants to the United States, Chiswick and Millers (2005) found a significant positive association between L2 English proficiency and the participants' educational level. They also found that L2 English proficiency is greater for individuals who immigrated at a young age and the longer the duration of residence. In addition, coming from a former US or British colony was an advantage, as was *not* having a refugee background (p. 27). Age at migration, educational attainment, and duration of residence were also significant factors for the proficiency level of L2 English and L2 French among male immigrants in Canada (Chiswick & Miller, 2001). In this study, a linguistic distance between the L1 and the L2 (English or French) was also associated with less use of the L2 and learner challenges (p. 405). The positive impact of exposure to the L2 before or after immigration was attested to in Raijman et al. (2015), who found that proficiency in Hebrew L2 increased with the length of stay in Israel, contacts with Israelis, and exposure to training in the Hebrew language (in the country of origin or in the host society; pp. 19–20). Finally, Ross (2000) studied to what extent several individual difference variables influenced the outcome of a language course programme preparing new migrants in Australia for English as a second language courses at a higher level. By using several different quantitative modelling techniques, he found that prior education and age of arrival were the most important predictors of achievement of competencies and certificate awards (p. 213). In addition, linguistic distance between L1 and English L2, hours of instruction, and length of residence in Australia were also significant in some of the statistical analyses.

As this brief summary has shown, the variables that have attracted some focus in existing research are educational background, residence, contact with L2 speakers, and age of arrival. The purpose of this study is to explore to what extent the hypothesised explanatory variables included explain variation in test outcomes. Although the current study does not investigate L2 learning as such, the dependent variable, the test result, may be perceived as a measure of learning outcome. Accordingly, the findings may also be relevant for the field of SLA.

Research question

This study aims to empirically identify the variables that affect adult immigrants' test results on a four-level scale based on the background information reported by

participants themselves. This following research question is addressed: *To what extent can prior education, region, years of residency, hours of L2 instruction, L1 language background, knowledge of English, L2 use, socialising with L1 speakers, age, and gender each explain variation in test results?*

Data

The materials in this study consist of test results and background information from 10,155 test takers who took *Norskprøve 2* (Norwegian Test 2) in 2009 and 2010. Norskprøve 2 was an official test of the Norwegian language for immigrants, which measured language at level A2 of the CEFR proficiency scale.[2] The test had an oral and written component.

The current analysis rests on test results from the three parts that comprise the written component—listening, reading, and writing. The three separate and autonomous parts were also assessed separately, and immigrants had to pass all three parts in order to pass the written component. The test was based on the CEFR and A2 level descriptions. The test was developed and administered by professional test developers at the Folkeuniversitetet/University of Bergen on behalf of official Norwegian authorities, and detailed test specifications were developed.[3] The test received the Q-mark from the Association of Language Testers (ALTE) in Europe in 2008, which means that the test has been thoroughly evaluated through a quality auditing system of European language tests operated by ALTE.[4]

The test result represents the dependent variable or the response variable that is expected to be influenced by the background or independent variables that the current study attempts to explain. The four levels of the variable are accounted for in Table 1, which provides information on how the 10,155 test takers are distributed according to their separate test results on the three subtests in the written component.

Table 1. Response variable (dependent variable).

Variable	Subcategory	*N*	Percentage
Test result	Passed 0 of 3 subtests	998	9.8
	Passed 1 of 3 subtests	1321	13.0
	Passed 2 of 3 subtests	1373	13.5
	Passed 3 of 3 subtests	6463	63.7

The distribution across the four categories of test outcomes reveals that the dependent variable is skewed, with a large majority of observations in the category *Passed three of three subtests*. In the statistical analysis, every possible combination of the passed categories is counted. Accordingly, there is an inherent assumption of unidimensionality built into the statistical model.

Participants

The test takers were all adult immigrants to Norway. They came from a heterogeneous group of learners comprised of a variety of social and educational backgrounds, L1 backgrounds and L2 proficiency levels, ages, and so on. Information about some demographic information was collected by a form requesting personal information that test takers completed when taking the test. This form was developed by those who develop and administer the language test, not by myself. The form collected information on L1, gender, age, home country, years of prior education, English proficiency, type and hours of instruction in Norwegian as the L2, length of residence in Norway, current employment status (working, studying, applying for a job), and L2 use outside the classroom. Some of these characteristics comprise the hypothesised explanatory variables (independent variables) in this study. However, the validity of some of the information must be considered because the questionnaire gathered self-reported data. For instance, it is unclear whether those test takers who self-identified as having an intermediate or advanced level of English truly demonstrated their claimed proficiency level. Furthermore, for the question regarding the degree of contact with the Norwegian language and Norwegians, the test takers were only given options such as *yes*, *no*, *never*, *seldom*, and *daily*, which are rather crude descriptive categories for capturing the complex nature of degree of contact with L1 speakers of Norwegian.

Another aspect that must be considered is related to the conditions for the collection of background information, which have consequences for the generalisability of the results. The test takers were asked to fill out the form on the day of the exam, but this was not mandatory. Moreover, the data in the current study come only from test takers who explicitly stated on the form their agreement that their test results and background information could be used for research purposes. Accordingly, the current sample is not random but must be described as a convenience sample. Although this makes it difficult to generalise the results to the whole population of immigrants that take the Norwegian language test, the current sample is large and comprises a substantial variation in backgrounds (see, for instance, Tables 2 and 3). Thus, the current data set can be seen as broadly representative of the heterogeneous immigrant population in Norway.

Table 2. Interval explanatory variables (independent variables).

Variable	N	No response	Mean	Standard deviation	Min.	Max.
Age	9503	652	31.8	9.0	15	70
Years of prior education	9568	587	12.6	3.9	0	27
Hours of L2 instruction	8274	1881	575.7	513.3	0	3300

Table 3. Nominal and ordinal explanatory variables (independent variables).

Variables and subgroups	N	Percentage
Germanic L1 background (nominal)		
Non-Germanic	9573	94.8
Germanic	522	5.2
No response	60	
Knowledge of English (ordinal)		
Beginner	1924	20.4
Basic	2858	30.4
Intermediate	3048	32.4
Advanced	1581	16.8
(L1 speaker)	323	
No response	421	
Region (nominal)		
Europe	2380	23.5
North America	120	1.2
Oceania	43	0.4
Latin America and the Caribbean	477	4.7
Asia	5162	50.9
Africa	1952	19.3
No response	21	
Residency (years, ordinal)		
>1 year	1638	16.2
>2 years	3394	33.5
>3 years	1986	19.6
>4 years	970	9.6
>5 years	606	6.0
More than 5 years	1525	15.1
No response	36	
L2 use (ordinal)		
Never	236	2.4
Seldom	2096	21.0
Weekly	1895	18.9
Daily	5771	57.7
No response	157	
Socialising with L1 speakers (ordinal)		
Never	1392	13.8
Seldom	3168	31.5
Weekly	2649	26.3
Daily	2858	28.4
No response	88	
Gender (nominal)		
Male	4010	39.5
Female	6144	60.5
No response	1	

The hypothesised explanatory variables of the study

The information on prior education, language (L1 and knowledge of English), L2 (instruction and use), and L1 community (residency and socialising with L1 speakers) yielded ten independent variables that represent the hypothesised explanatory variables in the current study. These are separated for analysis into three types: interval, nominal, and ordinal.

The three interval variables included in the analysis are age, years of prior education, and hours of L2 instruction, as given in Table 2.

The nominal and ordinal variables and their subcategories are listed in Table 3, which also provides information about the number of test takers in each category and the number of missing responses.

The variable *Germanic* warrants further explanation. It is well known from the SLA literature that congruence between L1 and L2 is an important constraint on L2 learning (Jarvis & Pavlenko, 2008). It is generally assumed that it is easier for someone to learn a language that is typologically related to one's own than to learn typologically distant languages; in Chiswick and Millers (2001), linguistic distance between the L1 and the L2 was among the significant predictors of L2 proficiency and use. Accordingly, it is important to include information about the typological distance of the learners' L1. However, the number of different L1s represented in the current data set (167) was too large to be exploited in analysis, and the number of speakers for each of the total number of 167 L1s also varies significantly (unfortunately, the test takers were not given the chance to list multiple L1s). Many languages, 105 out of 167, were spoken by fewer than 10 participants, and 23 of the languages were spoken by more than 150 participants. The largest L1 groups in the data set, with more than 500 speakers each, were Polish, Persian, Thai, and Arabic. The 167 L1s were categorised and broken down into 20 language families based on the top-level classification in *Glottolog 4.1*, a catalogue of the languages, language families, and dialects of the world developed by researchers at the Max Planck Institute for the Science of Human History (Hammarström et al., 2019). However, this typological information was not exploited in the analysis, because the 20 language families included several subgroups—broad categories encompassing languages that are linguistically quite diverse. For instance, many of the Indo-European languages are very different from Norwegian, although they technically belong to the same language family. Accordingly, information about which language family the test takers' L1s belong to is not precise enough to provide valuable information about the effects of typological proximity. Instead, the exploration of whether typological proximity influenced the test result is based on a binary variable created to separate learners with a Germanic language background from those with a non-Germanic one (Germanic/non-Germanic).

The variable *knowledge of English* is based on self-assessment, and it is interesting to explore if the learners' knowledge of English affected their performance in another Germanic L2, Norwegian. As the purpose of this variable was to explore the potential effects of knowledge of L2 English on L2 Norwegian, the 323 English L1-speaking participants were excluded from the statistical analysis. As it is important to assess the role of English language proficiency in the explanation of the test outcomes, and as the English L1-speaking participants only constituted 3.2% of the original data set, it is reasonable to exclude this group from the statistical analysis.

The variable *region* also requires further comment. The test takers were distributed across 154 countries and 6 major regions (Africa, Asia, Oceania, Europe, Latin America, and the Caribbean and North America[5]). However some languages, such as Bulgarian, Haka Chin, and Estonian, are reported as L1 only by learners of the same nationality, many of the other languages are spoken by learners who come from multiple countries or even different regions. For instance, L1 speakers of Arabic reported 21 different home countries in Africa and Asia. Although a few cases show a one-to-one relationship between L1 background and home country, a variable of *region* may supply non-linguistic information that the linguistic variable Germanic cannot do by emphasising precisely which part of the world the learners come from. Because educational levels, average income, and popular occupations vary across the different regions of the world, such a variable would also carry important information about the socio-economic status associated with each home country region. For instance, the United Nations (UN) Department of Economic and Social Affairs categorises countries and regions into "more developed/developed" and "less developed/developing" classifications based on information about generalised socio-economic conditions (note that these groups are "intended for statistical purposes and do not express a judgement about the stage reached by a particular country or area in the development process"; United Nations, 2022b). Countries in Europe and North America are characterised as "more developed" and countries in Africa, Asia (except Japan, where 16 of the learners in the current data set come from) and Latin America as "less developed." Furthermore, the countries in the region of Oceania, such as Australia and New Zealand, are regarded as "more developed," except Melanesia, Micronesia, and Polynesia, which are characterised as "less developed" (only seven of the learners in this study come from these subregions of Oceania). There is doubtless vast variation in the socio-economic backgrounds of the learners who come from the same world region. Still, this variable will be used as an indication of socio-economic conditions because it is relevant to explore whether the home region influences the learning outcome due to major social and economic differences among the six major regions that exist at the group level. These regions not only differ in the level of educational opportunity, average income, and common occupations of the population that lie at the heart of the UN's categorisation of countries, but significant cultural differences among the regions must also be considered when interpreting the results of the analysis.

The values of the other non-interval variables, *knowledge of English, residency, L2 use, socialising* (with L1 speakers), and *gender*, were the same in this analysis as the response options on the background questionnaire given to the participants. The variables L2 use and socialising were formulated from the answers to the questions: How often do you speak Norwegian outside the classroom? How often do you get together with Norwegian friends?

Statistical measures

The data were analysed using an ordered logistic regression. Regression is a method of investigating the relationship between a set of independent variables (predictor variables), which, in this study, was composed of the background information reported by the test takers themselves, and a dependent or response variable, the test result. As the dependent variable in the current study was ordinal (ranked, with four possible outcomes), an ordered

logistic regression was employed using the statistical software package Stata 16. Logistic regression analyses the probability of an event's occurrence, for instance, passing no part, some or all parts of a test. Other assumptions apply to logistic regression than to the more well-known and more widely used linear regression.[7] Logistic regression does not require a linear relationship between the dependent and independent variables but requires linearity of independent variables and log odds. Ordered logistic regression assumes that the observations are independent of each other, with minimal or no multicollinearity among the independent variables. Moreover, logistic regression usually requires large sample sizes.

In this study, the important question was not how much of the variance in the test result could be explained by the predictor variables but *if* the variables displayed statistically significant estimated effects on the dependent variable—the test result. The background information cannot be expected to explain a large part of the variation in the test results, as the assessment is solely based on the test takers' performance on the written component (listening, reading, writing). Rather, the intention was to explore whether a set of background variables may influence the test outcome.

In the first step of the model building, the 10 predictor variables were simultaneously entered into the model. The second step involved removing variables that (1) did not meet the statistical test criteria, even after attempts to rectify these issues, or (2) did not have any significant effect on the analysis. The final model (Appendix 1) was the result of some back and forth between the possible models, as some of the variables did not pass the required statistical tests or had to be square- or cube-root transformed because not all the variables had the necessary distribution for the logistical/ordinal regression.

To investigate the effects of the variables used in the final model, the margins command in Stata was used to calculate predicted probabilities (UCLA Statistical Consulting Group [UCLASCG], 2020; Williams, 2012). This is easier to understand than the odds ratio, which is commonly used to estimate the strength between an independent (predictor) variable and a dependent (response) variable.

Results

An ordered logistic regression (ologit command in Stata) was performed with 10 independent variables. A Brant test was used to test the parallel lines assumption, and it did not hold. The variable *Germanic* was identified as being problematic, and the reason is most likely that it had very few observations for one of the values (522 vs 9573). In such cases, the recommendation is to use a generalised ordered logit model with partial proportional odds (gologit2 programme), which "can fit models that are less restrictive than the parallel-lines models fitted by ologit (whose assumptions are often violated) but more parsimonious and interpretable than those fitted by a nonordinal method, such as multinomial logistic regression" (Williams, 2006, p. 58). Although the variable *L2 use* turned out to be significant in the ologit analysis, it was excluded from the analysis as its presence led to basic test assumptions being violated. This is because the variable was highly skewed with very few observations in one of the values (*never*, see Table 3). The ologit analysis also revealed that coming from two of the six world *regions* had statistically significant effects on the test result: Asia ($Z = 11.11, p > |z| = .000$) and Africa ($Z = 11.14, p > |z| = .000$).

However, as this variable was also highly skewed with very few observations in three of the values (North America, Oceania, Latin America, and the Caribbean, see Table 3), it could not be included in the analysis. In this case, the solution was not to exclude it entirely from further analyses, but instead to transform it into a binary collapsed variable with two values, which separated test takers into two groups: those coming from a European country and those originating from a country outside Europe.

Accordingly, nine dependent variables were included in the ordered logistic regression analysis conducted by means of a generalised ordered logit model (gologit2). Appendix 1 presents the Stata output from this analysis, which was tested for specification error (link test, $Prob > F = 0.0000$) and multicollinearity ($VIF = 1.23$). The reported pseudo-R^2 value for the model is .09.[8]

The generalised ordered logit model revealed that eight variables significantly impacted the dependent variable (test result) (Appendix 1).[9] Five of the significant variables had positive estimated effects on the test result, and three variables negatively influenced the test outcome (Appendix 1). Having a *Germanic* L1 background had a positive effect on the test outcome for the analyses of the probabilities of passing two ($Z=2.96$, $p > |z| = .003$) or three exam parts ($Z=2.96$, $p > |z| = .000$). *Knowledge of English* beyond the beginner and basic level also influenced the probability, and the Z-values in Appendix 1 reveal that the positive association between knowledge of English and the test result was slightly tighter in the analysis of the probability of passing all three exam parts ($Z=11.52$, $p > |z| = .000$). The variable *residency* also positively impacted the test outcome; however, the impact is only significant if the number of years living in Norway exceeds five, and moreover, only in the analyses of the probability of passing two ($Z=5.67$, $p > |z| = .000$) or three exam parts ($Z=5.77$, $p > |z| = .000$). There was also a positive estimated effect on test outcome on years of prior *education*, and this effect was significant ($p > |z| = .000$) across all categories of test outcomes, although the Z-values are slightly decreased in the last analysis ($Z=7.92$, 11.47, 10.83). Being a *female* test taker also positively influenced test result, and the effect increased slightly across the test outcome categories (Z-values ranging from 4.62 to 6.29). The variable *region* negatively impacted the test outcome across all categories, with negative Z-values increasing from −5.63 to −10.93, and the variable age followed the same pattern (Z-values ranging from −5.80 to 12.32). Finally, the number of hours of L2 instruction displayed a negative relationship with the test scores, once all other variables were accounted for. This effect was significant ($p > |z| = .000$) across all categories of test outcomes, with a slight decrease in the Z-values for the last analysis ($Z=-4.48$, −6.00, −5.15).

Table 4 shows what the results in the Stata output of Appendix 1 look like in terms of estimated probabilities of the distribution of test results based on the regression analysis. The listed probabilities are calculated using the margins command in Stata (UCLASCG, 2020). This approach is used to make the results more "tangible" and to get a "practical feel" for the impact of the significant variables in a logistic regression model (Williams, 2012, pp. 308–309). For the ordinal variables, the estimate is based on the values that the variables can take. For the interval data, the estimate is based on the mean and the standard deviation (*SD*) (plus/minus one *SD*) or the mean and the 25th and 75th percentile. In interpreting the results from the regression model and the estimated probabilities in Table

Table 4. Predicted probabilities for four possible test outcomes.

Variable	Value	Estimate			
		Pass 0/3 parts	Pass 1/3 parts	Pass 2/3 parts	Pass 3/3 parts
Germanic or non-Germanic L1	Non-Germanic	0.06	0.12	0.14	0.68
	Germanic	0.00	0.01	0.10	0.89
English	Beginner	0.08	0.17	0.18	0.57
	Basic	0.06	0.15	0.17	0.62
	Intermediate	0.04	0.11	0.13	0.72
	Advanced	0.02	0.07	0.09	0.82
L2 instruction	245	0.04	0.11	0.14	0.71
	400	0.04	0.12	0.15	0.69
	750	0.05	0.14	0.15	0.66
Region collapsed	Europe	0.02	0.07	0.10	0.81
	Outside Europe	0.05	0.15	0.16	0.64
Residency	<1 year	0.05	0.15	0.16	0.64
	>2 years	0.05	0.14	0.15	0.66
	>3 years	0.05	0.14	0.14	0.67
	>4 years	0.04	0.12	0.13	0.71
	>5 years	0.04	0.10	0.16	0.70
	more than 5 years	0.03	0.07	0.12	0.78
Education	10	0.06	0.15	0.16	0.63
	12	0.05	0.13	0.15	0.67
	16	0.03	0.09	0.13	0.75
Gender	Male	0.06	0.15	0.16	0.63
	Female	0.04	0.11	0.14	0.71
Age	23 (22.8)	0.03	0.09	0.11	0.77
	32 (31.8)	0.04	0.12	0.15	0.69
	40 (40.7)	0.06	0.16	0.18	0.60

4, it is important to bear in mind that the estimates show the relationship "on an 'all other things equal' basis" (Williams, 2012, p. 311).

A general observation in Table 4 is that the model captures the differences between test takers who passed all exam parts and those who failed one or more. For all the variables, the differences in percentage are notable in the column for the estimated probability of passing all three exam parts. The largest difference in the estimated distribution is found between test takers with Germanic L1s and test takers with non-Germanic L1s; test takers with Germanic L1s had a 21% higher probability of passing all parts of the test (89%) compared with test takers without (68%). The probability of passing all three exam parts gradually increased with more advanced English skills (from 57% to 82%). Test takers who come from a European country had a 17% higher probability of passing all parts of the test (81%) compared to test takers coming from outside Europe (64%). As for residency, test takers who had lived in Norway for more than five years have a 14% higher probability of passing all parts of the test (78%) compared to test takers who have

been in the country for less than a year (64%). For the probability of passing the whole test, a difference of a few years seems to matter. Test takers with 16 years of prior education had an 8% higher probability of passing the whole exam compared to learners with 12 years of prior education, and 12% higher probability compared to learners with 10 years of education from their home country. Test takers in their early twenties had a higher probability of passing all parts of the test (77%) compared to older test takers. The number of hours of L2 instruction and gender, although significant variables in the model, affected the probability of obtaining one of the possible test outcomes to a lesser degree than the other variables.

Discussion

This paper aimed to determine the role of background variables in the understanding of variations in learning outcomes and test performances among adult L2 learners. Based on a large data set of 10,155 immigrants who took an official language test in Norwegian and supplied information about their background on the day of the test, this study explored to what extent *prior education, region, years of residency, hours of L2 instruction, L1 language background, knowledge of English, L2 use, socialising with L1 speakers, age*, and *gender* could explain the variations in the test results. This research question was explored using ordered logistic regression. This revealed that (1) eight of the background variables, which represented the hypothesised explanatory variables, significantly impacted the dependent variable (test result) and (2) the effects of the background variables were not graded; they affected the probability of passing one, two, or all three subtests or not passing any.

The analysis showed that having a Germanic L1 background positively impacted the probability of passing all parts of the test. In addition, knowledge of another Germanic language, English, also significantly increased the probability of passing all three subtests. Accordingly, *typological proximity*, or more precisely, the typological distance between L1 and L2, seemed to influence the learners' chances of passing one, two, or all three parts of the L2 test. In consideration of the heavily researched phenomenon of crosslinguistic influence (CLI) in L2 acquisition, which identifies the degree of congruence between the L1 and the L2 as the most central constraint for CLI (Jarvis & Pavlenko, 2008), the facilitating effect of having a Germanic L1 background was to be expected. L1/L2 typological distance was also a significant factor in Ross's (2000) study of individual difference variables that influenced learning outcomes among recent migrants to Australia (see literature review). However, based on knowledge of how educational attainment and enrolment vary worldwide, a correlation between prior education and L1 Germanic was *als*o expected. Accordingly, the effect of L1 proximity on the current study should not *only* be interpreted as a reflection of the role that the typological relationship between the L1 and L2 plays in L2 acquisition but also interpreted as the observation possibly being partly explained by the variable of prior education.

The results of the current study showed that *residency* positively affected test outcomes, but the effect was significant only after 5 years of residence in Norway. This is consistent with previous research that showed that a long period of residence in the host country is a facilitative variable for L2 learning success (e.g. Chiswick & Miller, 2001, 2005; Raijman et al., 2015; Ross, 2000). The participants in the current study, as well as the studies described in the literature review, were all post-critical period learners or late/

adult L2 learners (e.g. Abrahamasson et al., 2018). In accordance with other studies on migrant L2 learning success, age acts as a constraint on performance (and learning outcomes), although the variable in this analysis (similar to Ross, 2000) was the age at the time of taking the test rather than the age on arrival, as in, for instance, the studies of Chiswick and Miller (2001, 2005) and Strube et al. (2013). Strube et al. also discussed the relationship between arrival age and length of residence, and, based on their findings and findings from related studies (i.e. Baynham et al., 2007; Condelli et al., 2003; Kurvers & Stockmann, 2009), these results "indicate that the younger learner has an advantage over the older learner, which is not compensated by a longer LOR [length of residence]" (2013, p. 61).

Across theoretical perspectives in the field of SLA, *exposure to L2 input* is considered one of the most important components in the process of learning an L2. Formal language training in a classroom setting is a significant source of data for L2 learning, and for many learners, the classroom is their only source of exposure to the language. However, migrant L2 learners potentially also have access to input outside the classroom, as the L2 is the language of communication in the host society. The amount of classroom instruction and the degree to which immigrants have contact with L2 speakers affect the amount of L2 exposure and accordingly act as sources of individuals' varied learning outcomes. The analysis showed that the variables in the current study related to L2 exposure pointed in different directions. The *socialising* variable was not significant in the statistical analyses, and the *L2 use* variable was deleted from the final analysis (using a generalised ordered logit model) because its presence led to the violation of basic test assumptions. However, *L2 use* was significant in the initial ordered logistic regression. The *L2 instruction* variable measured the amount of formal exposure to the L2, but in the current study, it had a significant *negative* effect on the test results. This contradicted Raijman et al.'s (2015, pp. 19–20) finding that proficiency in Hebrew L2 increased with contact with Israelis and exposure to training in the Hebrew language (in the country of origin or the host society). Moreover, Raijman et al.'s result was consistent with Baynham et al. (2007), who found a positive association between the number of classroom hours per week and L2 performance (and to some extent, Ross, 2000, p. 209). The results of Kurvers and Stockmann (2009) and Condelli et al. (2003) are however in accordance with the finding of the current study; they attested a negative correlation between the number of classroom hours and L2 reading skills. Strube et al. (2013) also investigated formal exposure (classroom hours), but their results were not significant. Accordingly, findings concerning the effect of the amount of formal instruction are inconclusive.

It was, however, difficult to interpret the *L2 instruction* variable in the current study. The data were provided by the participants, and it is reasonable to assume that they reported the number of hours of instruction they were eligible for as participants of official and state-funded language training programmes for immigrants (according to the Integration Act [IA], 2020, Ministry of Labour and Social Inclusion). The number of hours of instruction varied according to the participants' backgrounds, and an immigrant's educational background is the most important criterion for determining learning goals and how much language training they will receive (Ministry of Labour and Social Inclusion [IA, 2020]). Learners with little or no prior education are offered more classroom hours than learners with an academic background. In other words, in the same way that the observation that L1 Germanic languages lead to greater odds of passing all

subtests can be partly explained by prior education, the negative association between L2 instruction and test outcomes may also be interpreted in the context of the other significant variables. The impact of prior education on test outcomes, in particular, and the positive effect of having a Germanic L1 background and emigrating from a European country (across all categories of test outcomes, coming from a country outside Europe influenced the test result negatively) are important for understanding the role L2 instruction plays in the model. Learners with a Germanic L1 background are most likely to originate from a European country, and the populations of many European countries tend to be highly educated. Consequently, it may be that the role of the L2 instruction variable in the model reflected these circumstances rather than the role that L2 instruction in and of itself had on the test outcomes.

This possibility is relevant in the light of the findings of previous studies on the role of another measure of formal L2 exposure, *rate of attendance.* Kurvers and Stockmann (2009), Baynham et al. (2007), and Condelli et al. (2003) detected a positive effect of attendance on L2 learning. Although their analyses also produced an non-significant result for the attendance variable, Strube et al. (2013) concluded that "attendance is a crucial factor for learning" and "even more important than [the] number of scheduled classroom hours" (p. 62). There may have been another possible interpretation for the negative association between L2 instruction and test outcomes in the current study. This would imply that there exists an upper limit for the number of classroom hours (e.g. Kurvers, 2007), as participants that had been offered a maximum number of hours of classroom instruction did not seem to benefit from it. As suggested by Strube et al., it could be that the programme structure (e.g. intensity and regularity) matters more for L2 progress and L2 outcomes, at least for L2 learners with no or little schooling from their home country and/or with low levels of literacy, known as LESLLA learners.

The participants in the current study took their L2 test before language requirements for citizenship and permanent residency were introduced in Norway for the first time in 2017[10] and a few years before it became compulsory to sit for a language test after having completed a specific number of hours of training on Norwegian language, which was introduced in 2013. Accordingly, the consequences of failing the test for the participants in the current study were far less severe than the consequences that adult immigrant test takers currently face in Norway. Yet, the roles of the background variables explored in the current study to explain the variation in test outcomes are still relevant today. Such empirical findings are of value when discussing the consequences of the formal language requirements for residency and citizenship that have been introduced in a growing number of European states, Australia, and the USA. The current analysis indicated that the test takers who participated in the current study who had a background and profile that corresponded to low-literacy/low-education L2 learners, the LESLLA learners, and those who had not come from WEIRD countries (see the Literature Review section) were most likely to fail some or all of the language tests. In the literature, it has been increasingly acknowledged that LESLLA learners face particular challenges in adult language learning (Kurvers et al., 2015) and when taking language tests (Carlsen, 2017). As stated in Rocca et al.'s (2020) survey *of language and knowledge of social policies for migrants* in Europe, the set language requirements often exceed what is reasonable to expect from learners with limited education (p. 70). The identification and exploration of the factors that make it easier for some learners to learn a language or hinder or make it more

challenging for others are critical to unearthing empirical evidence that sheds light on the political aspects of language requirements and identifies the potentially discriminatory effects of language tests. As underscored by Shohamy and McNamara (2009), ever since the introduction of obligatory language and knowledge-of-society tests for migrants applying for residency and citizenship worldwide, language testers have been raising questions about "the use of tests within different political and social contexts" (p. 1). The concept of fairness in testing and the social and ethical dimensions of language testing are important not only for current professional language testers but also for the education of future language testers. Studies such as the current one offer important empirical evidence that there are personal test taker characteristics, often related to the test taker's background, that impact test performance and must be considered when constructing tests involving language abilities.

The results of the current study also have implications for the SLA research field and the education of future L2 teachers. The findings suggest that SLA students and future SLA teachers should be made (even more) aware of the importance of L2 learners' backgrounds and experiences and consequences these may have for their progress in learning the L2 and the outcome of the language learning. Several of the background variables in the current study were related to the environments surrounding the learners, particularly pre-immigration conditions, such as educational opportunities earlier in life. Some learner groups are less equipped for the language learning process because of their previous experiences and the conditions under which they have lived. Of course, this is in addition to the more learner-internal factors—traditional individual difference (ID) factors, such as language aptitude and motivation—that also create differences among adult refugees learning an L2. Thus, introductory books for SLA should include, to a larger extent, not only the typical ID factors that primarily relate to the individual learner but also the role of background factors in L2 learning. Motivation is a typical ID factor that is also of great importance for understanding adult L2 learning, yet such internal factors must be understood in relation to previous experiences and surrounding environments. Regardless of how motivated learners are, having received little or no education in their home country puts them in a difficult position, and the resulting path towards a functional level in the L2 is long.

Concluding remarks

This study aimed to explore whether the background information supplied by immigrants who took an official language test in Norwegian affected their chances of passing one, two, or all three parts of the test (three subtests in the written component of an official language test in Norwegian for immigrants). The most significant finding was the opportunity to statistically document that test takers' backgrounds affected the outcome of the official language test that measured L2 proficiency. The analysis revealed that eight variables significantly impacted the test results: prior education, region (location of their home country), L1 background, knowledge of English, L2 instruction, years of residence, age, and gender. Several aspects of the current study need to be noted. First, the study used self-reporting as the main data-gathering technique. The background information, analysed as predictors of the test result, was reported by the learners. Second, the study's approach was purely quantitative and based on a large group of

individual learners. Therefore, the profiles of the individual learners were concealed, and the generalisations drawn from the regression analysis are not relevant for every single adult learner of Norwegian L2. This study was explorative, and more research is needed. Third, the reported findings specifically addressed outcomes from the non-oral subtests test component, and the extent to which these findings contributed to the modelling of the oral component of the test is a highly interesting but still open question. However, some of the significant variables in the current study were also significant in Strube et al.'s (2013) study of oral development. Nonetheless, individual L2 learners' starting points in the learning process varied tremendously, and the current study demonstrated that these different starting points matter.

Although the impact of background variables was investigated in the Norwegian context, the findings are likely to hold across immigrant communities re-settling in other countries. Several of the significant background variables were related to pre-immigration conditions, such as educational opportunities earlier in life, and this underscores the importance of variables other than those that have traditionally been studied in individual differences research. Hence, the findings are also relevant for the study of variations in L2 learning outcomes. Adult L2 learning departs from multiple settings. This study contributes to the understanding of the role of migrant learners' backgrounds in shaping outcomes for L2 tests and therefore the lives of migrants settling into new communities.

ORCID iD

Ann-Kristin Helland Gujord https://orcid.org/0000-0002-5916-5323

Notes

1. https://www.leslla.org/
2. Norskprøve 2 and Norskprøve 3 do not exist any longer as they were replaced in 2014 with one language test, *Norskprøven* (Norwegian language test), which measures language proficiency at CEFR levels A1–B2. https://www.kompetansenorge.no/tests/norwegian-language-test/about-the-test/#ob=25753.
3. These are confidential.

4. https://www.alte.org/Setting-Standards.
5. I relied on the United Nations' geoscheme for this division of the world into regions and sub-regions (United Nations, 2022a).
6. For assumptions on logistic regression, see, for example, Liao (1994).
7. Stata reports McFadden's R^2 when referring to Pseudo R^2 (Williams, 2020).
8. The variable socialising is not significant.
9. Citizenship language requirements are currently B1 oral (raised from A2 oral in 2021) and A1 oral for permanent residency.

References

Abrahamasson, N., Hyltenstam, K., & Bylund, E. (2018). Age effects on language acquisition, retention and loss. Key hypothesis and findings. In K. Hyltenstam, I. Bartning & L. Fant (Eds.), *High-level language proficiency in second language and multilingual contexts* (pp. 16–48). Cambridge University Press.

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

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.

Baynham, M., Roberts, C., Cooke, M., Simpson, J., Ananiadou, K., Callaghan, J., McGoldrick, J., & Wallace, C. (2007). *ESOL effective teaching and learning* (ESOL Main Report 17/1/07). National Research and Development Centre for Adult Literacy and Numeracy).

Carlsen, C. H. (2017). Giving LESLLA learners a fair chance in testing. In M. Sosiński (Ed.), *Literacy education and second language learning by adults: Research, policy and practice* (pp. 135–148). University of Granada.

Chiswick, B. R., & Miller, P. W. (2001). A model of destination-language acquisition: Application to male immigrants in Canada. *Demography*, *38*(3), 391–409. https://doi.org/10.1353/dem.2001.0025

Chiswick, B. R., & Miller, P. W. (2005). Do enclaves matter in immigrant adjustment?. *City & Community*, *4*(1), 5–35. https://doi.org/10.1111%2Fj.1535-6841.2005.00101.x

Citizenships Act. (2020). Lov om endringer i statsborgerloven (krav om ferdigheter i norsk muntlig). (Act on changes in the Citizenships Act (proficiency requirements in oral Norwegian). Ministry of Education and Research. Retrieved from https://lovdata.no/dokument/LTI/lov/2020-10-30-123

Condelli, L., Wrigley, H. S., Yoon, K., Cronen, S., & Seburn, M. (2003). *What works: Study for adult ESL literacy students: Final report*. American Institutes for Research.

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

Darvin, R., & Norton, B. (2014). Social class, identity, and migrant students. *Journal of Language, Identity & Education*, *13*(2), 111–117. https://doi.org/10.1080/15348458.2014.901823

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing*. Cambridge University Press.

Dewaele, J. M. (2009). Individual differences in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (2nd ed., pp. 623–646). Emerald Group.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Lawrence Erlbaum.

EU-OECD. (2016). *How are refugees faring on the labour market in Europe? First evaluation based on the 2014 EU labour force survey ad hoc module* (DG EMPL Working Paper 1/2016).

Extra, G., Spotti, M., & Van Avermaet, P. (2009). *Language testing, migration and citizenship: Cross-national perspectives on integration regimes*. Continuum.

Hammarström, H., Forkel, R., & Haspelmath, M. (2019). *Glottolog 4.1*. Max Planck Institute for the Science of Human History.

Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and Brain Sciences*, *33*, 61–83. https://doi.org/10.1017/S0140525X0999152X

Hogan-Brun, G., Mar-Molinero, C., & Stevenson, P. (2009). *Discourses on language and integration: Critical perspectives on language testing regimes in Europe*. John Benjamins.

Integration Act. (2020). *Lov om integrering gjennom opplæring, utdanning og arbeid (Act on integration through training, education, and labour)*. Ministry of Labour and Social Inclusion. Retrieved from https://lovdata.no/dokument/NL/lov/2020-11-06-127

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. Routledge.

Kozar, O., & Yates, L. (2019). Factors in language learning after 40: Insights from a longitudinal study. *International Review of Applied Linguistics in Language Teaching*, *57*(2), 181–204. https://doi.org/10.1515/iral-2015-0113

Kunnan, A. J. (1995). *Test taker characteristics and test performance: A structural modeling approach*. Cambridge University Press.

Kurvers, J. (2007). Development of word recognition skills of adult L2 beginning readers. In N. Faux (Ed.), *Low-educated second language and literacy acquisition. Proceedings of the second annual forum* (pp. 23–43). Literacy Institute at Virginia Commonwealth University.

Kurvers, J., & Stockmann, W. (2009). *Alfabetisering NT2 in beeld: Leerlast en succesfactoren* [Focus on L2 literacy: Study load and success factors]. University of Tilburg.

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). CLS.

Liao, T. F. (1994). *Interpreting probability models: Logit, probit, and other generalized linear models*. SAGE.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–104). American Council on Education and Macmillan.

Norton, B., & Toohey, K. (2001). Changing perspectives on good language learners. *TESOL Quarterly*, *35*(2), 307–322. https://doi.org/10.2307/3587650

OECD. (2019). *International migration outlook 2019*.

Ortega, L. (2005). For what and for whom is our research? The ethical as transformative lens in instructed SLA. *Modern Language Journal*, *89*(3), 427–433. https://doi.org/10.1111/j.1540-4781.2005.00315.x

O'Sullivan, B., & Green, A. (2011). Test taker characteristics. In L. Taylor (Ed.), *Examining speaking: Research and practice in examining second language speaking* (pp. 36–64). Studies in language testing 30. UCLES/Cambridge University Press.

Raijman, R., Semyonov, M., & Geffen, R. (2015). Language proficiency among post-1990 immigrants in Israel. *Journal of Ethnic and Migration Studies*, *41*(8), 1347–1371. https://doi.org/10.1080/1369183X.2014.982523

Rocca, L., Carlsen, C. H., & Deygers, B. (2020). *The 2018 Council of Europe and ALTE survey on language and knowledge of society policies for migrants*. Council of Europe & ALTE.

Ross, S. J. (2000). Individual differences in learning outcomes in the Certificates in Spoken and Written English. In G. Bridley (Ed.), *Studies in immigrant English language assessment* (Vol. 1, pp. 191–214). National Centre for English Language Teaching and Research.

Shohamy, E., & McNamara, T. (2009). Language tests for citizenship, immigration, and Asylum. *Language Assessment Quarterly*, *6*, 1–5. https://doi.org/10.1080/15434300802606440

Strube, S., van de Craats, I., & van Hout, R. (2013). Grappling with the oral skills: The learning process of the low-educated adult second language and literacy learner. *Apples—Journal of Applied Language Studies*, *7*(1), 45–65.

Tarone, E., & Bigelow, M. (2005). Impact of literacy on oral language processing: Implications for second language acquisition research. *Annual Review of Applied Linguistics*, *25*, 77–79. https://doi.org/10.1017/S0267190505000048

UCLA Statistical Consulting Group. (2020). *Ordered logistics regression.* https://stats.idre.ucla.edu/stata/dae/ordered-logistic-regression/

United Nations. (2022a, April 7). *Methodology.* https://unstats.un.org/unsd/methodology/m49/

United Nations. (2022b, April 7). *Definition of regions.* https://population.un.org/wpp/DefinitionOfRegions/

van de Craats, I., Kurvers, J., & Young-Scholten, M. (2006). Research on low-educated second language and literacy acquisition. In I. van de Craats, J. Kurvers & M. Young-Scholten (Eds.), *Low-educated second language and literacy acquisition: Proceedings of the Inaugural Symposium Tilburg 2005* (Vol. 6, pp. 7–23). LOT.

Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *The Stata Journal, 6*(1), 58–82.

Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *The Stata Journal, 12*(2), 308–331. https://doi.org/10.1177/1536867X1201200209

Williams, R. (2020). *Scalar measure of fit: Pseudo R2 and information measures* (AIC & BIC). https://www3.nd.edu/~rwilliam/stats3/L05.pdf

Young-Scholten, M. (2013). Low-educated immigrants and the social relevance of second language acquisition research. *Second Language Research, 29*(4), 441–454. https://www.jstor.org/stable/43103934

Appendix 1

Generalised Ordered Logit Estimates				Number of obs = 6732			
				LR chi2(42) = 1168.76			
				Prob > chi2 = .0000			
Log likelihood = −6250.9729				Pseudo-R2 = .0855			
CollapsedGrade~g	Coef.	Std. Err.	z	p > |z|	[95%	Conf.	Interval]
0							
GERMANIC ENGLISH	15.67765	1051.394	0.01	.988	−2045.017		2076.373
1	.2313125	.1081746	2.14	.032	.0192942		.4433308
2	.7732993	.1272623	6.08	.000	.5238699		1.022729
3	1.691653	.2422211	6.98	.000	1.216909		2.166398
L2INSTRUCTION_SQ	−.0214392	.0047841	−4.48	.000	−.030816		−.0120625
REGION RESIDENCY	−.8741735	.1553572	−5.63	.000	−1.178668		−.569679
1	−.0686331	.1619002	−0.42	.672	−.3859517		.2486856
2	−.0891708	.176358	−0.51	.613	−.4348261		.2564845
3	.0447755	.2074078	0.22	.829	−.3617364		.4512874
4	.2322625	.2348553	0.99	.323	−.2280455		.6925705
5	.3397885	.1934375	1.76	.079	−.039342		.7189191
EDUCATION	.1025248	.0129464	7.92	.000	.0771502		.1278993
GENDER	.4155117	.0898633	4.62	.000	.2393828		.5916405

(continued)

Appendix 1. (Continued)

Generalised Ordered Logit Estimates				Number of obs = 6732			
				LR chi2(42) = 1168.76			
				Prob > chi2 = .0000			
Log likelihood = −6250.9729				Pseudo-R2 = .0855			
CollapsedGrade~g	Coef.	Std. Err.	z	p > |z|	[95%	Conf.	Interval]
AGE	−.0317519	.0054701	−5.80	.000	−.0424732		−.0210306
_cons	2.657682	.2821993	9.42	.000	2.104582		3.210783
1							
GERMANIC	2.985859	1.009653	2.96	.003	1.006976		4.964743
ENGLISH							
1	.1673878	.0814256	2.06	.040	.0077965		.326979
2	.6662078	.0897017	7.43	.000	.4903957		.8420198
3	1.274652	.1368135	9.32	.000	1.006502		1.542801
L2INSTRUCTION_SQ	−.0210966	.0035164	−6.00	.000	−.0279886		−.0142045
REGION	−.9201104	.101094	−9.10	.000	−1.118251		−.7219698
RESIDENCY							
1	.0529484	.1030365	0.51	.607	−.1489995		.2548963
2	.0364537	.1144354	0.32	.750	−.1878357		.260743
3	.2424736	.1387443	1.75	.081	−.0294601		.5144074
4	.3457461	.1579563	2.19	.029	.0361575		.6553347
5	.7497457	.1323022	5.67	.000	.4904381		1.009053
EDUCATION	.1067248	.009304	11.47	.000	.0884892		.1249604
GENDER	.3572578	.0644309	5.54	.000	.2309755		.4835401
AGE	−.0398809	.003917	−10.18	.000	−.0475581		−.0322037
_cons	1.710579	.1995459	8.57	.000	1.319477		2.101682
2							
GERMANIC	1.245346	.318852	3.91	.000	.6204074		1.870284
ENGLISH							
1	.1966177	.0754964	2.60	.009	.0486474		.3445879
2	.6856237	.0798778	8.58	.000	.529066		.8421814
3	1.264586	.1097917	11.52	.000	1.049398		1.479774
L2INSTRUCTION_SQ	−.0167415	.0032503	−5.15	.000	−.0231119		−.0103711
REGION	−.8810416	.0805862	−10.93	.000	−1.038988		−.7230954
RESIDENCY							
1	.0790108	.0857088	0.92	.357	−.0889754		.246997
2	.1024823	.0971677	1.05	.292	−.0879628		.2929274
3	.2936735	.1191023	2.47	.014	.0602373		.5271098
4	.2230696	.1364706	1.63	.102	−.0444079		.4905472
5	.6442214	.1116415	5.77	.000	.4254079		.8630348
EDUCATION	.0911145	.0084168	10.83	.000	.0746179		.107611
GENDER	.3645703	.0570396	6.39	.000	.2527748		.4763659
AGE	−.0439194	.0035642	−12.32	.000	−.0509051		−.0369337
_cons	1.111832	.1767455	6.29	.000	.7654177		1.458247