

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Tammelin-Laine, Taina; Hirvelä, Tuija; Huhta, Ari; Neittaanmäki, Reeta; Ohranen, Sari; Stordell, Elina

Title: Predicting Placement Accuracy and Language Outcomes in Immigrants' L2 Finnish Education

Year: 2018

Version: Published version

Copyright: © Georgetown University Press, 2018.

Rights: In Copyright

Rights url: <http://rightsstatements.org/page/InC/1.0/?language=en>

Please cite the original version:

Tammelin-Laine, T., Hirvelä, T., Huhta, A., Neittaanmäki, R., Ohranen, S., & Stordell, E. (2018). Predicting Placement Accuracy and Language Outcomes in Immigrants' L2 Finnish Education. In J. M. Davis, J. M. Norris, M. E. Malone, T. H. McKay, & Y.-A. Son (Eds.), *Useful Assessment and Evaluation in Language Education* (pp. 149-165). Georgetown University Press.

Useful Assessment and Evaluation in Language Education

Editors

John McE. Davis, Yorktown Systems Group

John M. Norris, Educational Testing Service

Margaret E. Malone, Georgetown University

Todd H. McKay, Georgetown University

Young-A Son, Georgetown University

GEORGETOWN UNIVERSITY PRESS

Washington, DC

© 2018 Georgetown University Press. All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and recording, or by any information storage and retrieval system, without permission in writing from the publisher.

The publisher is not responsible for third-party websites or their content. URL links were active at time of publication.

Library of Congress Cataloging-in-Publication Data

Names: Georgetown University Round Table on Languages and Linguistics (2016 : Washington, D.C.), author. | Davis, John McE., 1970- editor.

Title: Useful Assessment and Evaluation in Language Education / editors, John McE. Davis, John M. Norris, Margaret E. Malone, Todd H. McKay, Young-A Son.

Description: Washington, D.C. : Georgetown University Press, 2018. | Series: The Georgetown University Round Table on Languages and Linguistics series | Includes bibliographical references and index.

Identifiers: LCCN 2017019313 (print) | LCCN 2017032351 (ebook) | ISBN 9781626165400 (ebook) ISBN 9781626165397 q(hc : qalk. paper) | ISBN 9781626165403 q(pb : qalk. paper)

Subjects: LCSH: Language and languages--Study and teaching--Evaluation--Congresses.

Classification: LCC P53.63 (ebook) | LCC P53.63 .G46 2016 (print) | DDC 418.0071--dc23

LC record available at <https://lccn.loc.gov/2017019313>

∞ This book is printed on acid-free paper meeting the requirements of the American National Standard for Permanence in Paper for Printed Library Materials.

19 18 9 8 7 6 5 4 3 2 First printing

Printed in the United States of America

Chapter 9

◆ Predicting Placement Accuracy and Language Outcomes in Immigrants' L2 Finnish Education

TAINA TAMMELIN-LAINE, ARI HUHTA, REETA NEITTAANMÄKI, TUIJA HIRVELÄ, and SARI OHRANEN

Centre for Applied Language Studies, University of Jyväskylä, Finland

ELINA STORDELL

Testipiste, Helsinki, Finland

ASSESSMENTS USED IN INTEGRATION training can be divided into three types or purposes: placement, formative, and summative (see, e.g., Brown 2012). The current study focuses on placement assessment although data from the final summative assessments also contribute to the investigation.

The purpose of the placement assessment described in this study is to guide immigrants to the most suitable training track with respect to their readiness for studies and the most suitable language module according to their Finnish language proficiency (Finnish National Board of Education 2012a). Correct placement is important in terms of time and other resources, but it also saves both immigrants and their teachers from frustration resulting from placement in an inappropriate level or type of course, even if it is possible for learners to change courses or tracks if need be.

Placement of immigrants in training tracks is not standardized. A national curriculum (Finnish National Board of Education 2012a) provides only very broad guidelines for integration training. Decisions about placement and other types of assessment (such as formative and summative), as well as about instruction, are made locally at the level of regional employment areas and individual institutions. The national curriculum, however, defines the target level for second language (L2; Finnish or Swedish) studies at the end of integration training. Placement is typically based on combining several sources of information, including immigrants' initial

Finnish language proficiency and information considered to indicate their readiness for formal language studies, such as their previous educational level and whether they have studied languages before.

Recently, a particular approach to placement was promoted in a project at the Centre for Applied Language Studies (CALS) at the University of Jyväskylä with funding from the Ministry of Economic Affairs and Employment, which is responsible for immigrants' integration training. The ministry cannot dictate that a specific placement procedure be used across the country because the administrators of both placement assessment and integration training are selected on the basis of competitive tendering in different regions. The ministry was, however, concerned about the comparability of placement procedures and was, therefore, funding a project that aimed to create and maintain a network of institutions that will commit to using the same procedure. The placement procedure designed by Testipiste, an organization specializing in immigrants' placement assessment, was selected by the ministry as the system to be advanced at the national level.

Starting systematic research on the validity of the Testipiste placement system was one of the aims of the project. Informal feedback gathered from the users of the Testipiste placement system has been quite positive. However, more empirical evidence about the procedure is needed.

Two issues have been addressed in our research so far and are reported here. The first concerns the lack of precise guidelines for assessors to combine different kinds of information collected during placement; recommendations about the most suitable track are based on assessors' subjective judgment about the importance of different pieces of information but also on feedback received from the teachers. To begin to formulate guidelines for combining and possibly weighing different kinds of information in placement recommendations, thus improving their comparability across assessors and institutions, we examined which information contributed to assessors' decisions.

The second issue we investigated was the extent to which the placement procedures predict immigrants' attainment of L2 proficiency, which is the main target in integration training. Effective placement assessment "should reflect the features of the teaching context" (Davies et al. 1999, 145). At this stage of research, the only information available to us about the teaching context was the track in which the learners had participated; therefore, we began our research by investigating the relationship between placement assessments and the (language) outcomes of teaching. In the future, we plan to examine the predictive validity of placements by gathering detailed information about the characteristics and activities in the different training tracks, as will be discussed at the end of this chapter.

There is no previous research on how placement assessments relate to outcomes in the context of training of immigrants in Finland; such studies also appear rarely internationally (however, see Gonzalves 2016). However, understanding how placement procedures relate to success, or lack thereof, in training is potentially useful information both for improving placement procedures and for increasing the use of placement information by the teachers (e.g., for identifying learners who might struggle if left without additional support).

Our results indicate that only certain types of information collected during placement are used by assessors for making placement recommendations and that only some of that information predicts L2 learning outcomes. Besides improving our understanding of the placement procedure, such findings can be used to shorten the procedure by removing uninformative parts from it, thus making it more practical. This research can also help us improve current placement tasks and develop new procedures.

Immigration and Language Learning in Finland

In 2015, approximately 6% of the population of Finland spoke other native languages in addition to Finnish, Swedish, or Sami, which are the traditionally spoken languages of the country (Statistics Finland 2016). The largest language groups were Russian (22% of all with a foreign background), Estonian (14.6%), Somali (5.4%), English (5.4%), and Arabic (5.1%; Statistics Finland 2016). Because of the growing number of immigrants, the Act on the Promotion of Immigrant Integration (1386/2010) came into effect in 2011 to (1) support immigrant integration; (2) make it easier for immigrants to play an active role in Finnish society; and (3) promote gender equality, nondiscrimination, and positive interaction between different population groups. Integration training is considered key to becoming a member of Finnish society since it includes learning the Finnish or Swedish language and communication skills,¹ learning civic and working-life skills, and participating in guidance counseling (Finnish National Board of Education 2012a).

Integration training in Finland is divided into two separate types of courses: one for adults who are literate in any language and another for nonliterate adults. Both are implemented to enhance professional, job-related adult education. Instruction is full time, based on the national curriculum, and led by professional teachers. The training is cost free and learners receive integration assistance and compensation for expenses for the course days. The length of education depends on immigrants' previous skills and needs, which are evaluated during an interview as part of the placement assessment. The number of students in integration training in 2015 was 14,742 (Työministeriö 2016).

For literate adults, the maximum length for integration training is sixty credit units (two thousand one hundred lesson hours during approximately one year), and the goal of language training is level B1.1 (i.e., low B1 on the Common European Framework of Reference [CEFR]; Finnish National Board of Education 2012a). This level is a threshold for applying for many benefits, including Finnish citizenship or entering many vocational training programs. The types of available integration-training courses are divided into different tracks that differ by speed or pace of instruction: slow, intermediate, and fast.

Nonliterate adults first participate in literacy training (a maximum of forty credit units). The objective is to learn basic oral and written Finnish or Swedish skills and to attain, on average, A1.2; that is, mid-A1 on the CEFR (Finnish National Board of Education 2012b). However, according to Tammelin-Laine (2014), many adults with no previous education do not achieve the targeted Finnish skills during

literacy training in order to continue on to integration training. The acquisition of literacy skills is particularly challenging for them.

Placement Assessment as a Decision-Making Instrument

Placement assessment is widely used at colleges and universities, for example, to divide students into homogeneous groups based on their language abilities (Green and Weir 2004). Plakans and Burke (2013) argue that the potentially high stakes of placing students into different program levels calls for a careful understanding of test use, decision-making, and the impact of test results on test-takers' lives. Therefore, the use of standardized proficiency tests for placement purposes, while quite common in some contexts, can be problematic (see Kokhan 2013) since, for instance, such tests might not be able to address the needs of the particular context. Placement of immigrants into language training is probably a context in which language proficiency test scores alone are not sufficient, as other types of information about immigrants' backgrounds are likely needed for appropriate placement decisions, particularly if training courses differ not only in terms of their language requirement but also, for example, in their pace of progress.

We are not aware of any international surveys of placement assessment of immigrants, but it is likely that most countries with language programs for immigrants use procedures that are designed, or adapted, for the specific context. Such procedures probably vary considerably according to country. Canada, for example, uses the Canadian Language Benchmarks Placement Test (CLBPT), referenced to the Canadian Language Benchmarks (Hajer and Kaskens 2012). CLBPT is a task- and competency-based standardized assessment tool testing L2 English skills for communicating in the real world. In contrast, the German placement-assessment system monitors learners' motivation and other indicators of learning progression alongside language skills (Perlmann-Balme and Dengler 2007).

Testipiste: Finnish Placement Assessment for Immigrant Adults

The current study relates to a project titled "Finnish Placement Assessment for Immigrant Adults" (2015–16), coordinated by CALS at the University of Jyväskylä and funded by the Ministry of Economic Affairs and Employment. The project focused on a placement-assessment system for L2 Finnish developed in 2010–13 by Testipiste, an assessment center for adult migrants, originally funded by the European Social Fund. The project aimed to improve placement assessment in Finland, to standardize it by encouraging a wider use of the Testipiste system, and to develop the system further. Research reported here contributes to the further development of the system.

Testipiste placement test procedures start with an interview (thirty minutes), which includes word dictation and mechanical reading. If the interview reveals participants have low literacy skills, they continue with more detailed literacy tests. If

they have some oral skills in Finnish, they will continue with the speaking test (fifteen minutes long). Next, most participants will take tests on morphological reasoning and basic mathematics (fifteen minutes each). If they have some oral and written Finnish ability, examinees proceed to the reading (one hour), listening (one hour), and writing tasks (forty minutes).

The recommendations given by the assessor(s) are used at employment services for placing the migrants into integration training. All the tracks except the literacy track are divided into four modules according to the starting level of language proficiency (see table 9.1). For the literacy track, the starting level of language modules is based on literacy skills. The tracks differ slightly in terms of their goals and L2 proficiency starting levels; the main difference between the tracks is the pace of study.

Placement tests cannot fully predict learners' course performance because of the effect of many contextual variables (Green and Weir 2004, 474). However, placement assessment gives information that is useful for teachers by showing, for example, what kind of tasks the participants are used to working with, and whether, as is the case in the partly computerized Testipiste system, they are familiar with using a computer.

◆ Table 9.1. Language module CEFR levels available in capital region in each integration-training track

Language module/ Starting level	Literacy track	Slow track	Intermediate track	Fast track
1	Below 0	0	0	0
2	...	A1.3	A1.3+	A2.1
3	...	A2.1	A2.1+	A2.2
4	...	A2.2	A2.2+	B1.1
Goal	A1.2	A2.2–B1.1	B1.1	B1.1–B1.2

Methods

The objective of the study was to improve our understanding of the placement procedure developed at Testipiste by (1) investigating how different types of information contributed to assessors' recommendations about placement of immigrants in training tracks, and by (2) examining to what extent placement could predict L2 learning outcomes at the end of integration training. While the placement of immigrants into appropriate levels of language modules within each training track is a key aspect of the placement process as a whole and was also examined in the current study, we do not cover that in detail here. The main reason for this is the fact that the use of L2 test results in the placement of immigrants into language modules is very straightforward; almost everybody was placed into the level indicated by their lowest result across speaking, listening, reading, and writing. In contrast, placement into training tracks and prediction of L2 learning outcomes is much more complex and therefore deserves to be analyzed in more detail.

Given these concerns, the current study had the following two research questions:

- RQ1. What information determines the subjective placement of the learners into the training track (literacy, slow, intermediate, or fast)?
- RQ2. Do the different parts of the placement assessments explain the L2 proficiency achieved by the immigrants at the end of the integration training?

Participants

Two hundred eighty-six immigrants participated in the study. They came from (a) five institutes of adult education and (b) one private company providing integration training in the capital region. Altogether, seventy-one nationalities with sixty native languages were represented among the participants, including Russian (21.7%) and Arabic (11.5%), which are also among the five biggest language groups in Finland (Statistics Finland 2016). Most participants were female (69.6%; male: 30.4%). Participants' mean age was 33 years ($SD = 8$; under 30: 41.6%; 30–39 years: 39.2%; over 40: 19.2%). The mean length of residence in Finland before the placement assessment was twenty-two months ($SD = 31.7$). However, 50% of the participants had taken the placement assessment within seven months after their arrival to Finland. Participants' educational background varied considerably; 23.4% had a maximum of nine years of previous formal education; 4.2% reported having no education at all; 33.9% had graduated from vocational education or an upper-secondary school; and 42.7% had a bachelor's degree, polytechnic diploma, or a master's degree. The participants had studied 1.1 languages on average (excluding Finnish; $SD = .8$), but they reported knowing 1.4 languages on average ($SD = .8$).

All four tracks were included in the recommendations for the immigrants in our study (literacy track: 11.2%; slow track: 15.0%; intermediate track: 69.6%; fast track: 4.2%), but in practice, they studied only in the intermediate or slow track, apparently because integration training in the fast track program was not available for them in the particular institutions. Those who received a literacy track recommendation attended literacy training first and then continued to integration training on the slow track. It is important to note that track recommendations in the current sample differ notably from recommendations for all immigrant learners in the capital region as whole. In 2015, based on a total of 3,868 examinees, the percentages of the track recommendations were as follows: literacy 19.2%, slow 24.9%, intermediate 44.8%, and fast 8.8%.

Data Collection and Analysis

Data were collected from 286 immigrants who (a) participated in the placement assessments at Testipiste, (b) completed integration training, and (c) received final, summative grades in L2 Finnish in 2015–16.

A variety of assessment data-collection tools were used both in the placement assessment and the final, summative assessment to investigate the learners' language and other skills, as well as background information. All the language test tasks used in

the placement assessment were carefully developed and piloted with over a thousand learners, and cut scores for proficiency levels have been defined via standard-setting procedures.

Placement Assessments

In the Testipiste model, oral language skills are assessed during the placement assessment interview and in a separate speaking-test task integrated in the interview. In the interview, the participants are asked about the following background information topics in Finnish: name, address, phone number, country of origin, native language, age, time of arrival to Finland, previous Finnish or Swedish courses, known and studied languages, impression of themselves as language learners, length of education in home country, occupation and work experience, IT skills, motivation and capability to study Finnish or Swedish at school and at home, and plans and wishes for the future.

Participants with no or very little command of Finnish are interviewed primarily in a shared language or in their first language with the help of an interpreter. In the speaking-test task, the participants are first asked to describe a picture; the theme expressed in the picture is then discussed more widely and at a more general level, if possible. Both the interview and the speaking-test task are used for assessing the participants' speaking skills, and the need to take additional language tests is determined by that assessment.

The placement assessments were comprised of different components capturing four different sets of skills: (1) readiness, (2) writing, (3) reading, and (4) listening comprehension. The "readiness skills" tests aim to help estimate if the learner is ready to fully participate in formal education and possesses some of the basic study skills needed in integration training. These skills are assessed with the following four tests: word dictation, mechanical reading, morphological reasoning, and basic mathematics. Word dictation and mechanical reading are included in the interview part of the placement assessment procedure. Word dictation includes ten words. The first five are shared with all the participants. Then the last five words are selected either from the lower or higher level based on participants' performance in the first part. This task is used for assessing phonological working memory, knowledge of the Roman alphabet, understanding of grapheme-phoneme correspondences, and discrimination of Finnish phonemes. The mechanical reading (reading aloud) test is used for investigating participants' accuracy, fluency, and speed of reading a Finnish text—or their ability to read a text written in the Roman alphabet in the first place. It is not used for assessing reading comprehension or pronunciation. The test on morphological reasoning tests accuracy and fluency of reading, detection of similarities and differences in the elements of an artificial language, and making analogous conclusions based on the models resembling linguistic structures. The basic mathematics test includes such fundamental mathematical operations as addition, subtraction, multiplication, and division with whole numbers and decimals; percentage calculation; time transformations; and simple equations.

For assessing immigrants' writing skills in Finnish, there are tests at two level ranges. The lower level covers CEFR levels A1.3–A2.2, and the higher one covers

levels A2.1–B1.1. The decision on the appropriate level for the participant is made during the interview part. At both levels, tasks include writing about personal life and responding to an e-mail message. Additional tasks include picture-based writing (lower level) and expressing an opinion (higher level).

The reading and listening comprehension subtests include six to eight tasks with thirty to thirty-five multiple-choice or true/false items. The tests are administered at two difficulty levels in the same way as the writing tests. The texts are short messages and narratives, and the audio recordings are announcements, discussions, and interviews.

Final Summative Assessment at the End of Integration Training

In the institutions from which the current data come, the final assessment of learners' Finnish language skills at the end of integration training was conducted by the teachers with the help of an end-of-program test designed at Testipiste. However, the teachers combine the test results with the information they gather during the training by using a range of approaches, which varies among teachers. The final summative grades are, thus, not arrived at in a standardized way and obviously vary in terms of their reliability. Final language grades are expressed as CEFR levels (using the more fine-grade Finnish version of the CEFR scale) and reported separately for speaking, listening, reading, and writing.

Analyses

The contribution of different types of placement information to the recommendation of the training track (RQ 1) was investigated with an ordered probit regression analysis (in Mplus 7.4; Muthén and Muthén 2015) with the track as the dependent variable and age, number of studied languages, mechanical reading, word dictation, basic mathematics, morphological reasoning, length of residence in Finland, level of education, and gender as independent variables. The model was estimated by using a robust weighted least squares estimator. Ordered probit regression was used because the dependent variable (track recommendation) was an ordinal-scale variable. The 264 participants were analyzed in ordinal probit regression analysis. Of the 264, 5% ($n = 13$) were recommended for track 1 (literacy track); 16% ($n = 42$) for track 2 (slow track); 75% ($n = 198$) for track 3 (intermediate track); and 4% ($n = 11$) for track 4 (fast track). Some of the participants recommended for the literacy track could not be included in the analysis because they were not given all the "readiness" tests.

The assessors do not use a fixed formula for weighing specified factors for track recommendations. They are instructed to use the results of skills other than L2 Finnish (because L2 results are used for determining the starting level of the language course [module], not the track) and all relevant background information about the learner gathered during the placement interview. However, how the assessor balances all those factors is left to their judgment. Therefore, the current study was a post hoc analysis to discover which factors the trained and highly experienced assessors at Testipiste took into account when making training-track recommendations.

The relationship between placement assessment and final, summative language assessment (RQ 2) was investigated with a linear regression analysis with Mplus,

using MLR estimation (maximum likelihood parameter estimation, which is robust to nonnormality in the data). The assumptions underlying the use of linear regression (e.g., collinearity and distribution of residuals) were checked with IBM SPSS Statistics 22. The four language grades (speaking, listening, reading, and writing) were used as dependent variables and the same potentially predictive factors that were used in the analysis of the training-track recommendations were used as independent variables. In addition, some further variables were used as independent variables: those that turned out not to explain training-track recommendation (e.g., length of residence in the country) or that are not used in the placement for the track but rather in decisions about the L2 starting level (in this case, the level of speaking skills in Finnish that is assessed for all test-takers). The reason for including initial L2 speaking skills as one of the predictors was the possibility that differences in initial L2 proficiency persist till the end of training, despite teachers' efforts to bring immigrants' language skills to the same level.

Results

Results of Placement and Final Summative Assessments

Table 9.2 shows participants' performance on the four tests of "readiness" skills. Participants received the highest mean scores on word dictation and mechanical reading (83% and 88% of the maximum, respectively). The average results for basic mathematics and morphological reasoning were somewhat lower (65% and 72% of the maximum, respectively).

Table 9.3 presents learners' mean writing, listening, reading, and speaking scores from the placement assessments. Most (about 75%–80%) had such low Finnish skills that they participated in the speaking test only. Their mean speaking level was only slightly higher than A1.2, while the highest level achieved was B1.2. About one fourth

◆ Table 9.2. Results of the tests of readiness skills (placement assessment)

Placement assessment	N	\bar{X}	SD	Med.	Min.	Max.
Word dictation	285	2.54	0.65	3	0	3
Basic mathematics	265	13.15	5.17	14	0	20
Morphological reasoning	267	21.64	7.93	24	0	30
Mechanical reading	286	2.63	0.68	3	0	3

◆ Table 9.3. Proficiency in language skills (placement assessment)

Placement assessment	N	\bar{X}	SD	Med.	Min.	Max.
Speaking	286	2.25	2.85	0	0	12
Writing	71	5.24	1.76	5	2	10
Reading comp.	58	6.45	1.76	7	3	11
Listening comp.	56	6.68	1.98	7	3	11

Note: The results refer to CEFR levels; 0 = below A1.1 and 12 = B1.2.

of the test-takers participated in the writing test and one fifth in reading and listening assessments. In those tests, the mean level was A1.3–A2.1, and the highest level was above B1.1.

As table 9.4 shows, the speaking results were the highest and writing and reading comprehension had the lowest mean scores on the final, summative assessment after integration training. The lowest CEFR level attained was A1.2 and the highest was B2.1. The median level was A2.2.

◆ **Table 9.4.** Proficiency in language skills (final, summative assessment), $N = 286$

Final grade	\bar{X}	SD	Med.	Min.	Max.
Speaking	5.44	0.94	5	2	8
Writing	5.09	0.98	5	2	8
Reading comp.	5.09	0.88	5	3	8
Listening comp.	5.16	0.86	5	3	8

Note: The results refer to CEFR levels; 2 = A1.2 and 8 = B2.1.

Understanding Decisions on Placement to Training Tracks

Recommendations based on placement assessment in fact concern two aspects of integration training: estimation of Finnish language proficiency (language level) and the pace at which the learner is expected to make progress in their studies (track). As mentioned earlier, we focus on the training track placements (RQ1), as the way the assessors weigh different kinds of information from the background interview and readiness tests is unknown and likely to be much less straightforward than the placement into the appropriate L2 level module (which is based on the weakest score across the four language skills).

Table 9.5 summarizes the results of the regression analysis with the training-track recommendation (four tracks) as the dependent variable and the most likely predictors that the assessors were using for their recommendations as the independent variables. It should be noted that none of the nonsignificant variables were removed from the model because, at this stage, we wanted to investigate the contribution of all potentially useful variables. A very high proportion of the decisions (91.2%) could be explained with the following factors: readiness skills (mechanical reading, word dictation, and morphological reasoning) and certain background information, such as number of studied languages (excluding Finnish) and educational background. Learners' age, gender, or length of residence in the country did not explain track recommendations significantly.

Because all the statistically significant regression coefficients are positive, an increase in the test scores, number of languages studied, level of previous education, and so on relates to a higher (i.e., faster) track recommendation and vice versa.

Finally, a comparison of the adjacent threshold estimates in table 9.5 supports the assumption that the assessors could distinguish between four different tracks with the help of placement information. This is indicated by the fact that the estimates for the thresholds between the different tracks (i.e., between tracks 1 and 2,

◆ Table 9.5. Ordered probit regression model for predicting training-track recommendation (standardized model results), $N = 264$

Variables	Est. B	SE	p-value
Age (in years)	0.057	0.059	0.334
Number of studied languages	0.153	0.051	0.003
Mechanical reading	0.303	0.067	0.001
Word dictation	0.295	0.076	0.001
Basic mathematics	0.163	0.093	0.081
Morphological reasoning	0.251	0.080	0.002
Length of residence in Finland	-0.040	0.059	0.497
Level of education	0.159	0.069	0.022
Gender	0.064	0.047	0.174
Thresholds			
Track 1	3.356	0.423	0.001
Track 2	4.588	0.451	0.001
Track 3	6.573	0.475	0.001
R ²	0.912	0.026	0.001

tracks 2 and 3, and tracks 3 and 4) differed significantly from each other (at the .001 level).

In a separate analysis with only the readiness tests as predictors of track recommendations, they jointly explained 85% of the variance, from which we can deduce that background information plays a smaller role in the placement than readiness test results.

Relationship between Placement Assessment and Final Summative Assessment

As the first step in trying to understand to what extent it might be possible to predict ultimate (language) achievement in integration training, we examined the relationship between placement assessment and the final language grades given by the teachers at the end of the typically close-to-one-year courses. These final, summative assessments are a combination of an external, final test designed by Testipiste and the teachers' own continuous assessment. Final language grades are reported separately for the four skills on the Finnish version of the CEFR scale. We used the four final language grades as dependent variables in linear regression analyses. As independent variables, we used the four tests of "readiness" skills and certain background variables (age, gender, educational level, length of residence, and number of languages studied), as they are generally considered potentially important factors in language learning or they had been found statistically significant in our analyses of the track placements reported for RQ 1 above. We also used the actual track the learners had been through (only two tracks were in fact available to them) and the speaking-test grade from the placement test as further independent variables. The other language tests could not be used, as only speaking was assessed for practically

all learners during placement; the other skills were not assessed for learners who did not know any Finnish in the placement stage.

Table 9.6 shows the results for predicting the final reading and writing grade (these models include also the nonsignificant variables). A total of 36% of the variance in the final reading grades and 38% in the writing grades were explainable on the basis of the information gathered about learners' during placement procedures. Exactly the same variables turned out to be significant predictors for both skills. The findings indicate that women and younger learners achieved higher levels of reading and writing than men and older learners. Furthermore, higher scores in the basic mathematics test and higher initial speaking skills were related to better performance in reading and writing at the end of the integration training.

The findings concerning speaking and listening are somewhat different although there were some similarities with written skills. As table 9.7 demonstrates, the amount of explained variance was lower for the final oral skills—about 29% for speaking and 25% for listening—than for the written skills. Only three statistically significant predictors of listening comprehension could be identified; namely, age, the score on the basic mathematics test, and the initial speaking proficiency in Finnish. We note here that these same variables also predicted writing skills, as reported above, and that they were also statistically significant predictors of the final speaking grades (see table 9.7). However, two further variables were found to

◆ **Table 9.6.** Linear regression model for predicting the final grade in writing and reading skills (standardized model results), $N = 264$

Variables	Writing final grade			Reading comp. final grade		
	Est. B	SE	p-value	Est. B	SE	p-value
Age (in years)	-0.219	0.067	0.001	-0.212	0.063	0.001
Number of studied languages	0.082	0.057	0.148	0.093	0.063	0.143
Mechanical reading	0.072	0.052	0.169	0.028	0.051	0.577
Word dictation	0.008	0.049	0.871	0.075	0.051	0.139
Basic mathematics	0.358	0.081	0.001	0.363	0.083	0.001
Morphological reasoning	0.004	0.100	0.966	-0.009	0.102	0.933
Length of residence in Finland	-0.076	0.052	0.144	-0.061	0.059	0.302
Level of education	0.028	0.061	0.646	-0.019	0.058	0.738
Gender	-0.162	0.047	0.001	-0.143	0.045	0.002
Placement assessment speaking	0.145	0.051	0.004	0.163	0.056	0.004
Actual training track	0.093	0.077	0.226	0.112	0.079	0.156
Intercept	4.666	0.471	0.001	5.221	0.467	0.001
R^2	0.378	0.047	0.001	0.364	0.048	0.001

account for variance in the final speaking grades: number of languages studied (the more the better) and gender (women did better).

Discussion and Conclusion

Summary of the Findings

The first part of the study aimed at understanding the decision-making of the assessors at Testipiste in terms of the factors they consider when recommending a particular pace of learning (i.e., track; RQ 1). Results showed that the tests of readiness skills (mechanical reading, in particular) were the key determiners of the recommendation although specific background variables (number of previously studied languages and educational background) also contributed to the decision. Fewer than 10% of the track recommendations could not be explained from the variables identified in this study; this result suggests that the assessors at Testipiste work systematically. Although their interrater reliability could not be estimated from the available data, the fact that the track recommendations were separable from each other suggests they assess fairly consistently.

In the second part of the study, we investigated whether different parts of the placement assessment predict the final language grades (RQ 2). As could perhaps be expected, most of the variance in the final language grades could not be explained

◆ Table 9.7. Linear regression model for predicting the final grade in oral skills (standardized model results), $N = 264$

Variables	Speaking final grade			Listening comp. final grade		
	Est. B	SE	p-value	Est. B	SE	p-value
Age (in years)	-0.285	0.066	0.001	-0.279	0.066	0.001
Number of studied languages	0.161	0.067	0.016	0.095	0.068	0.162
Mechanical reading	-0.009	0.059	0.878	0.008	0.065	0.900
Word dictation	0.072	0.055	0.188	0.047	0.054	0.389
Basic mathematics	0.180	0.083	0.029	0.306	0.088	0.001
Morphological reasoning	0.027	0.101	0.789	0.002	0.107	0.986
Length of residence in Finland	-0.009	0.061	0.879	-0.060	0.062	0.334
Level of education	-0.084	0.062	0.174	-0.055	0.065	0.393
Gender	-0.111	0.051	0.030	-0.088	0.051	0.082
Placement assessment speaking	0.226	0.052	0.001	0.196	0.063	0.002
Actual training track	0.131	0.090	0.145	0.044	0.077	0.568
Intercept	5.648	0.481	0.001	6.045	0.492	0.001
R ²	0.286	0.046	0.001	0.252	0.048	0.001

on the basis of placement assessments. However, for reading and writing, about one third of the variance could be predicted from previous performance on the tests of mathematics and speaking (Finnish), age, and gender. Prediction of listening and speaking was more modest, which suggests that progress in these may be more variable across individuals; it is probably affected also by the amount and nature of training at workplaces that immigrants have during integration training. Interestingly, initial differences in L2 speaking skills before starting the training seemed to persist to some degree up to the end of the program since command of spoken L2 predicted higher achievement in all four language domains. Perhaps even a small initial advantage in, for example, being able to follow instructions right from the start helps such learners to make more rapid progress compared with those with no or very little command of the L2.

Issues to Be Studied in the Future and Lessons Learned

The current study was a starting point for more extensive and longer-term research into the validity of the placement assessments used for immigrants' integration training in Finland, and more generally, into the effectiveness of the integration training system, particularly as regards immigrants' L2 learning. In the future, a network of institutions will use the Testipiste placement system and will engage in systematic study of the factors affecting learning outcomes. In the remainder of the chapter, we review the main issues with placement and final assessments, as well as integration training more generally, that we identified during the current study. We also discuss the types of data that will be needed in the future to study placement assessments as well as the effectiveness of integration training more thoroughly.

Current Placement Assessment System

Although we could explain over 90% of the variance in the track recommendations in an exploratory analysis of a number of variables, we still need to do further analyses (series of model fitting) to find out the most optimal combination of variables that explain placement recommendations. Determining if all the collected background information is relevant for track recommendations is rather straightforward and can be addressed, at least partly, with our current data. We know now which of the readiness tests and background information items are likely to contribute most to the recommendations, which paves the way for a construction of a formula for weighing the different factors that all assessors could use. This would increase the reliability of decisions, especially when new assessors are recruited. However, we would need a more comprehensive data set to validate the present findings because the literacy and fast track recommendations were clearly underrepresented in the data.

In the future, one of the foci will be research-based development of the readiness tests in cooperation with experts on special education in order to increase the validity of the tests. The measurement properties of some of these tests (word dictation and mechanical reading) might also be improved by lengthening their currently very short scoring scale.

In addition, more detailed information about the placement assessments and assessment processes than was available in the current study would also be useful in

future studies. We did not know which assessments were provided by which rater, nor was there detailed background information available about the raters. In addition, in some cases, two different versions of the tests were used, but information about a given version was not available; therefore, we could only assume that the scores provided from different test versions were comparable.

Final Summative Assessment

Although the same final language test was used in all the institutions from which data was gathered, the final language grades were affected by all the other information that the teacher had collected during the course. To enable a more precise evaluation of learners' L2 proficiency at the end of the training program, we need to know both the final test results and teachers' own evaluations based on continuous assessment of the learners. Currently, final assessment varies across institutions in the country, and only some use the tests designed at Testipiste. In future studies, we obviously need to make sure that data on learners' language achievement is gathered with the same, validated measures across the institutions involved in the research.

Effectiveness of Instruction in the Training Tracks

Placement assessment is an important part of immigrants' integration training system in Finland but only one part of the entire system. The current study was the first step in a larger-scale and longer-term investigation of the effectiveness of the integration training and, particularly, whether the division into four tracks functions as intended. Therefore, we conclude the discussion by outlining what future studies of the whole training system should consider.

Investigating the usefulness of track recommendations is a far more complex issue than can be captured by predicting those recommendations from information from the placement procedure. To study properly the meaningfulness of organizing training in terms of tracks requires that we know much more about how the tracks differ from each other in terms of teaching activities, materials, and methods used, as well as approaches to formative assessment and feedback. We know that slower tracks are typically longer in terms of duration and number of contact hours, which is one obvious way to try to ensure that both kinds of tracks reach the same goals. However, besides that, we know very little about how the tracks differ. Are the differences between different educational institutions and teachers salient enough that one can really identify a track? That is, are the between-track differences clearly bigger and more important than within-track differences? Therefore, one of the key requirements in the longer-term investigation of integration training is that we gather more comprehensive information about the training period itself.

More detailed information will be needed on both learners and teachers. We need, for example, to know the actual length of study for each learner, as well as information about their on-the-job training and L2 learning during that training. Furthermore, we should know about learners' motivation and attitudes, particularly toward using and learning the L2. As for teachers, information will be needed about their characteristics, such as teaching experience, preferred teaching and assessment methods, and strategies for providing learners feedback.

Conclusion

Increasing immigration to Finland poses challenges to L2 education, and one solution has been the creation of integration training programs with different tracks and language modules for different kinds of learners. We know, however, rather little about the effectiveness of integration training and how the different kinds of assessment (placement, formative, summative, etc.) function as part of the training system. To begin to address these gaps in our knowledge, the University of Jyväskylä and the Testipiste assessment center carried out a study, investigating the underlying characteristics of the track recommendations made in the placement process and the relationship between placement and final assessments. The study paves the way to more systematic and longer-term research on integration training in the future.

Note

1. According to the constitution of Finland, Finnish and Swedish are the two official languages of the country.

References

- Brown, Annie. 2012. "Uses of Language Assessments." In *The Encyclopedia of Applied Linguistics*, edited by Carol Chapelle. Chichester, Eng.: Wiley-Blackwell. doi:10.1002/9781405198431.wbeal1237.
- Davies, Alan, Annie Brown, Cathie Elder, Kathryn Hill, Tom Lumley, and Tim McNamara. 1999. *Dictionary of Language Testing*. Studies in Language Testing 7. Cambridge: Cambridge University Press.
- Finnish National Board of Education. 2012a. *National Core Curriculum for Integration Training for Adult Migrants 2012*. Helsinki: Finnish National Board of Education.
- . 2012b. *National Core Curriculum for Literacy Training for Adult Migrants 2012*. Helsinki: Finnish National Board of Education.
- Gonzalves, Lisa. 2016. "When Standardized Tests Fail: Informal Assessment of LESLLA Learners in California Adult Schools." Paper presented at the Low Educated Second Language and Literacy Acquisition (LESLLA) symposium, Granada, Spain, September 8–10, 2016. http://wdb.ugr.es/-leslla2016/wp-content/uploads/other-areas_otras-areas.pdf.
- Green, Anthony B., and Cyril J. Weir. 2004. "Can Placement Tests Inform Instructional Decisions?" *Language Testing* 21:467–94. doi:10.1191/0265532204lt2930a.
- Hajer, Anne, and Anne-Marie Kaskens. 2012. *Canadian Language Benchmarks: English as a Second Language for Adults*. Ottawa: Citizenship and Immigration.
- Kokhan, Kateryna. 2013. "An Argument against Using Standardized Test Scores for Placement of International Undergraduate Students in English as a Second Language (ESL) Courses." *Language Testing* 30:467–489. doi:10.1177/0265532213475782.
- Ministry of Employment and the Economy, Finland. 2011. Act on the Promotion of Immigrant Integration 1386/2010. Unofficial translation. www.finlex.fi/en/laki/kaannokset/2010/en20101386.pdf.
- Muthén, Linda K., and Bengt O. Muthén. 2015. *Mplus User's Guide*. 7th ed. Los Angeles: Muthén & Muthén. https://www.statmodel.com/download/usersguide/MplusUserGuideVer_7.pdf.
- Perlmann-Balme, Michaela, and Stefanie Dengler. 2007. *Einstufungssystem für die Integrationskurse in Deutschland. Handreichungen für Einstufende* [Placement assessment system for integration courses in Germany. Guidelines for placement]. Munich: Goethe-Institut.
- Plakans, Lia, and Maureen Burke. 2013. "The Decision-Making Process in Language Program Placement: Test and Nontest Factors Interacting in Context." *Language Assessment Quarterly* 10:115–134. doi:10.1080/15434303.2011.627598.

Statistics Finland. 2016. "Population 31.12.2015 by Year, Language, Sex and Age." http://pxnet2.stat.fi/PXWeb/pxweb/en/Maahanmuuttajat_ja_kotoutuminen/Maahanmuuttajat_ja_kotoutuminen_-_Maahanmuuttajat_ja_kotoutuminen/o3o_kieli.px/table/tableViewLayout1/?rxid=b81a4bfi-f54b-4969-a1ec-bd265db9d7bb.

Tammelin-Laine, Taina. 2014. *Aletaan alusta. Luku- ja kirjoitustaidottomat aikuiset uutta kieltä oppimassa* [Let's start from the beginning. Non-literate adults learning a new language]. PhD diss., University of Jyväskylä, Jyväskylä Studies in Humanities.

Työministeriö. 2016. *Työnvälitystilasto* [Statistics on employment services]. Helsinki, Fin.: The Ministry of Economic Affairs and Employment.