The Effects of Mobile-Assisted Language Learning on Language Acquisition and Integration Outcomes

Evidence from a Randomized Controlled Trial in Integration Courses in Germany

Zouhier Kassaballi

University of Erfurt

Abstract

This paper evaluates the effectiveness of Mobile-Assisted Language Learning (MALL) on language acquisition and integration outcomes among migrants and refugees in Germany. Using a randomized controlled trial conducted across integration courses in six German cities, I examine whether encouraging the use of language-learning applications improves German language proficiency and social integration. The intervention consisted of a motivational video featuring successful integration stories and informational flyers about recommended language-learning apps. Results from 317 participants across 22 classes indicate that the treatment significantly increased app usage: 46% of the treatment group downloaded the recommended "Ankommen" app compared to only 5% in the control group, and treatment participants spent on average 30 minutes daily on MALL activities compared to 15 minutes in the control group. Intention-to-treat estimates show that treatment assignment increased self-reported German language skills by 0.3 standard deviations. Complier average causal effect estimates suggest that actual MALL usage improves self-reported German skills by 0.78 to 0.83 standard deviations. These findings demonstrate that low-cost, scalable mobile learning interventions can effectively enhance language acquisition among migrants, with important implications for integration policy.

Keywords: Mobile-Assisted Language Learning, Migration, Integration, Language Acquisition, Randomized Controlled Trial, Germany

JEL Classification: J15, J61, I21, O33

1. Introduction

The successful integration of migrants into host countries has become one of the most pressing policy challenges facing developed economies. In Germany, this challenge has intensified following the large influx of asylum seekers between 2015 and 2018, during which approximately 5.5 million new migrants were registered (BAMF, 2019). Against the backdrop of demographic change and concerns about labor shortages, immigration represents a key policy lever for maintaining the labor force and economic prosperity (German Council of Economic Experts, 2022). However, the potential benefits of immigration can only be realized if migrants are successfully integrated into the labor market and society.

Language proficiency has been consistently identified as a crucial determinant of migrant integration. A substantial body of research demonstrates that host-country language skills are essential for labor market success (Aldashev et al., 2009; Bleakley & Chin, 2004; Chiswick & Miller, 2015; Dustmann & Fabbri, 2003; Schmid, 2023; Yao & Van Ours, 2015), educational achievement (Ruhose, 2013a, 2013b), and social and political integration (Bleakley & Chin, 2010). In recognition of this, the German government has established integration courses that provide language training alongside civic orientation. These courses are mandatory for refugees and asylum seekers, with 58% of the 256,238 mandated individuals participating in 2018 (BAMF, 2019).

Despite these policy efforts, outcomes remain concerning. In 2018, successful first-attempt completion rates for the final B1-level state examination were below 50%. This low success rate, combined with the difficulties migrants face in the labor market due to insufficient language skills, motivates the search for effective interventions to improve language learning outcomes. While the literature clearly establishes that language course participation improves language proficiency and subsequent integration outcomes (Foged et al., 2022, 2023, 2024; Foged & van der Werf, 2023; Heller & Slungaard Mumma, 2023; Kanas & Kosyakova, 2023; Lochmann et al., 2019; Schuss, 2018), relatively little is known about how to maximize the effectiveness of language learning.

This paper addresses this gap by experimentally evaluating the effectiveness of Mobile-Assisted Language Learning (MALL) among migrants and refugees in German integration courses. Language-learning applications such as Duolingo and similar platforms offer several attractive

features: they are user-friendly, low-cost, easily scalable, and highly flexible, enabling individualized learning paths and broad, barrier-free accessibility. I extend the traditional MALL approach by incorporating integration-related content and measuring both language and integration outcomes—what I term Mobile-Assisted Language Learning for Integration (MALLI).

The intervention design draws on insights from behavioral economics and educational psychology. I employ an encouragement treatment that includes a motivational video featuring successful integration stories of recent migrants who faced similar challenges. This approach leverages the concept of "near peer role models"—individuals who are similar to the target population—which has been shown to positively affect second language acquisition (Muir, 2018; Muir et al., 2021). The framing effect and the influence of positive future self-image are expected to change perceived benefits of integration and mobile learning (Sayman & Öncüler, 2008), while also addressing intertemporal choice considerations regarding investment in language learning.

The randomized controlled trial was conducted across integration courses at adult education centers (Volkshochschulen) in six German cities. Random assignment occurred at the class level, with a 50% treatment probability. The final sample includes 317 students from 22 classes, with 147 students in the treatment group and 185 in the control group. The intervention took place during the first week of each treatment class and included a seven-minute motivational video, informational flyers with an app portfolio and QR codes for easy download, and a question-and-answer session.

The results demonstrate that the intervention successfully increased MALL adoption. While only 5% of control group participants downloaded the recommended "Ankommen" app, 46% of treatment group participants did so. Treatment group members also reported spending significantly more time on language-learning apps—30 minutes per day compared to 15 minutes in the control group. These behavioral changes translated into improved outcomes: intention-to-treat estimates indicate that treatment assignment increased self-reported German language skills by 0.3 standard deviations. Complier average causal effect (CACE) estimates, using various definitions of compliance, suggest that actual MALL usage improves self-reported German skills by 0.78 to 0.83 standard deviations.

This study makes several contributions to the literature. First, it provides experimental evidence on the effectiveness of MALL for language acquisition among migrants—a population for which rigorous evidence has been notably lacking. While meta-analyses suggest MALL can be beneficial for second language acquisition (Burston, 2015; Hassan Taj et al., 2016), most existing evidence comes from studies of students in formal educational settings and focuses primarily on vocabulary acquisition rather than comprehensive language skills. Second, the study evaluates a low-cost, scalable intervention that could realistically be implemented at scale in integration policy. Third, the randomized design allows for credible causal inference, addressing concerns about selection bias that affect much of the quasi-experimental literature on language training effectiveness.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature on language acquisition, mobile learning, and migrant integration. Section 3 describes the institutional context of German integration courses. Section 4 details the experimental design and intervention. Section 5 presents the data and summary statistics. Section 6 describes the empirical methodology. Section 7 reports the results, and Section 8 concludes with policy implications and directions for future research.

2. Literature Review

2.1 Language Skills and Migrant Integration

The relationship between host-country language proficiency and migrant outcomes has been extensively studied across multiple disciplines. Economic research consistently demonstrates substantial wage returns to language skills. Dustmann and Fabbri (2003) find significant earnings premiums for English proficiency among immigrants in the United Kingdom. Bleakley and Chin (2004) use age at arrival as an instrument for language skills and find that English proficiency significantly increases earnings among childhood immigrants to the United States. Chiswick and Miller (2015) provide comprehensive evidence across multiple receiving countries, documenting the robust positive relationship between destination-language proficiency and labor market outcomes.

For Germany specifically, Aldashev et al. (2009) document substantial wage returns to German language skills among immigrants. Yao and Van Ours (2015) show that language proficiency affects not only wages but also employment probability. Schmid (2023) provides recent evidence

confirming the importance of German language skills for labor market integration of refugees. Beyond labor market outcomes, Ruhose (2013a, 2013b) demonstrates that language skills significantly affect educational achievement among immigrant children in Germany.

The effects of language proficiency extend beyond economic outcomes. Bleakley and Chin (2010) find that English proficiency promotes social and civic integration among immigrants in the United States, affecting measures such as social assimilation and political participation. This broader conception of integration aligns with policy objectives in Germany, where integration courses are designed to facilitate both labor market and social integration.

2.2 Language Training Programs

A growing body of research evaluates the effectiveness of language training programs for migrants. Lochmann et al. (2019) study the French language training requirement for immigrants and find that it significantly improves language proficiency and employment outcomes. Their quasi-experimental approach exploits policy variation to identify causal effects. Foged et al. (2022, 2023, 2024) provide extensive evidence from Denmark on the effects of language training, showing positive impacts on language skills and labor market integration. Foged and van der Werf (2023) examine the interaction between language training and other integration policies.

For Germany, Schuss (2018) evaluates the effectiveness of integration courses and finds positive effects on language acquisition. Kanas and Kosyakova (2023) examine language training among refugees in Germany and document improvements in language proficiency. However, most of this evidence relies on quasi-experimental methods, and the identifying assumptions may not always be satisfied. Heller and Slungaard Mumma (2023) provide a notable exception with their randomized evaluation of adult English language training in the United States, finding positive effects on language skills and employment.

2.3 Mobile-Assisted Language Learning

The rapid proliferation of smartphones has created new opportunities for technology-assisted learning. Mobile-Assisted Language Learning (MALL) refers to language learning that is facilitated or enhanced through the use of mobile devices such as smartphones and tablets. The

potential advantages of MALL include accessibility, flexibility, the ability to learn anywhere and anytime, and the possibility of personalized learning paths.

Burston (2015) provides a comprehensive meta-analysis of MALL project implementations over twenty years. The analysis reveals generally positive effects on language learning outcomes, though most studies focus on vocabulary acquisition. Sung et al. (2015) conduct a meta-analysis of mobile devices in language learning and find moderate positive effects. Cho et al. (2018) review the literature on mobile applications for language learning and document positive impacts on student engagement and learning outcomes.

Hassan Taj et al. (2016) conduct a meta-analysis specifically focused on MALL for second language acquisition. While they find evidence of positive effects, more than 60% of the studies examined rely on SMS-based learning methods, and only one study examines language skills beyond vocabulary. This limits the applicability of existing findings to modern app-based learning approaches. Critically, the evidence base for MALL effectiveness comes primarily from studies of school and university students, whose circumstances and motivations differ substantially from those of adult migrants learning a new language for survival and integration purposes.

2.4 Role Models and Motivation

The intervention design in this study incorporates insights from research on role models and motivation. Kosse et al. (2020) and Resnjanskij et al. (2024) demonstrate that pro-social role models and mentors can positively affect social behavior and economic outcomes among youth. Williamson et al. (2013) provide evidence on the effectiveness of mentoring programs for disadvantaged populations.

In the context of language learning, research in applied linguistics highlights the importance of motivation for successful acquisition. Dörnyei and Chan (2013) emphasize the role of possible selves and future self-guides in second language motivation. Thompson and Vásquez (2015) document how successful language learners can serve as motivational models for others. Particularly relevant is research on "near peer role models"—individuals who are similar to learners in relevant characteristics. Muir (2018) and Muir et al. (2021) show that such role models have positive effects on second language acquisition motivation and outcomes. These findings inform the design of the motivational video intervention used in this study.

3. Institutional Context: German Integration Courses

Germany's integration courses (Integrationskurse) were established by the Immigration Act of 2005 as the cornerstone of the country's integration policy. The courses are administered by the Federal Office for Migration and Refugees (BAMF) and delivered primarily through adult education centers (Volkshochschulen) as well as private language schools. The program is designed to provide migrants with basic German language skills and knowledge of German society, law, and culture.

The standard integration course comprises 600 hours of instruction divided into six modules of 100 hours each. Five modules (500 hours) are dedicated to language instruction, while one module (100 hours) covers civic orientation. The language component follows the Common European Framework of Reference for Languages (CEFR) progression, beginning at level A1 and targeting level B1 proficiency by the end of the course. Every two modules correspond to one CEFR level, with internal assessments conducted after the A1 and A2 levels and an official state examination at the course conclusion.

Participation in integration courses is mandatory for refugees and asylum seekers who do not already possess German language skills. In 2018, 58% of the 256,238 individuals mandated to attend actually enrolled in courses (BAMF, 2019). All new students undergo a mandatory placement test to ensure appropriate class assignment, which helps maintain comparable average language skills across courses. The courses follow a standardized BAMF curriculum, providing consistency in content delivery across different providers and locations.

The integration course system offers several advantages for conducting this research. First, the mandatory nature of participation for many refugees means the sample is not strongly pre-selected based on motivation or other unobservable characteristics that might bias results. Second, the standardized curriculum and placement testing ensure comparability across treatment and control groups. Third, the institutional setting allows for systematic intervention delivery and outcome measurement. Finally, the population of integration course participants represents the target group for potential policy scaling.

4. Experimental Design

4.1 Randomization Procedure

The experiment employed random assignment at the class level within A1 and A2-level integration courses at Volkshochschulen. Classes were randomized individually with a 50% probability of treatment assignment. This design allows for both treatment and control classes within the same institution, improving statistical efficiency while maintaining internal validity. The class-level randomization was necessary because the intervention is delivered to entire classes and to avoid contamination between treatment and control individuals within the same class.

The sampling frame consisted of all institutions registered on BAMF's public search engine within a specified geographic area centered on Göttingen. Participating institutions were located in Lower Saxony, Hessen, Bremen, and Hamburg—states that together hosted 23.1% of total integration course participants in 2018. Courses had different starting times, necessitating continuous enrollment into the study, with randomization occurring as each new class began.

4.2 Intervention

The intervention was an encouragement treatment designed to motivate and facilitate the use of language-learning applications as a supplement to regular integration course instruction. The treatment was administered during the first week of classes for all treatment groups and consisted of three components.

First, participants watched a seven-minute motivational video featuring stories of migrants and refugees who had recently arrived in Germany and successfully integrated despite facing similar challenges. The video was designed to induce a positive future self-image and leverage the concept of near peer role models. By showing individuals with similar backgrounds who had overcome obstacles through language learning and integration efforts, the video aimed to change perceived benefits of integration and mobile learning while addressing intertemporal choice considerations.

Second, participants received informational flyers presenting a portfolio of recommended language-learning applications. The flyer included QR codes and direct download links for easy app installation. The featured apps included Ankommen (an integration-focused app developed by German authorities), Deutsche Grammatik (a grammar training app), Memrise (a vocabulary

learning app), and Duolingo (a comprehensive language course app). Additional recommended apps from Deutsche Welle, the Goethe Institute, and Deutschakademie were also listed (See Figure 1).

Third, a question-and-answer session addressed participants' queries about the apps and mobile learning. This component ensured that participants understood how to access and use the recommended resources and could clarify any technical or practical concerns.

Figure 1: Intervention Flyer

4.3 Timeline

The study did not include a baseline survey; instead, randomization ensured comparability between treatment and control groups at the outset. The intervention was delivered during the first week of the module for all treatment classes. Data collection occurred at the end of the module (week 8 for A1 courses), with an endline survey administered by the researcher using paper questionnaires completed by students during class time (see Figure 2).

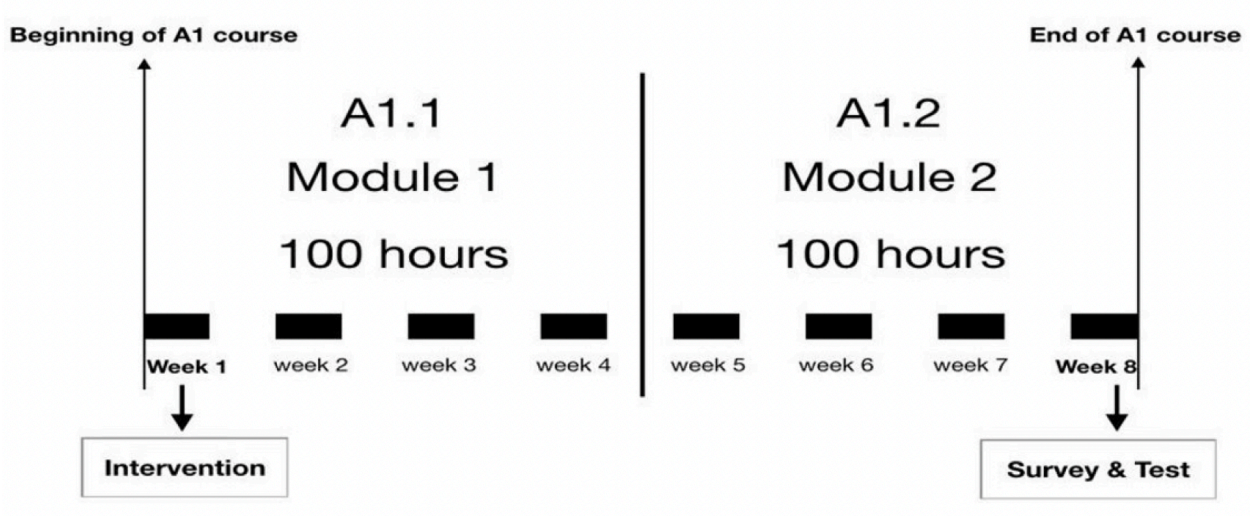

Figure 2: Study Design and Timeline *Notes:* The figure illustrates the timeline for A1-level courses, which span two modules (A1.1 and A1.2) of 100 hours each over approximately eight weeks. The intervention was administered during Week 1 of the course. The endline survey and language assessment were conducted during Week 8, at the conclusion of the A1 course level. The same structure applies to A2-level courses.

4.4 Pilot Study

A pilot study was conducted in 2019 across six integration courses at Volkshochschulen in Göttingen and Braunschweig, involving 81 students. Following Eldridge et al. (2015), the pilot served multiple purposes: testing the intervention's efficacy and estimating the direction of main outcomes; assessing participation and consent rates; evaluating research design feasibility; refining measurement instruments through in-depth interviews; and estimating parameters for the pre-analysis plan and power calculations. The pilot findings informed the design and implementation of the main study.

5. Data

5.1 Sample Description

The data come from three sources: the endline survey, teacher reports, and in-class test results. The final dataset comprises 317 students from 22 different classes across six cities: Bremen, Peine, Hannover, Braunschweig, Göttingen, and Kassel. The treatment group includes 147 students from 10 classes, while the control group includes 185 students from 12 classes.

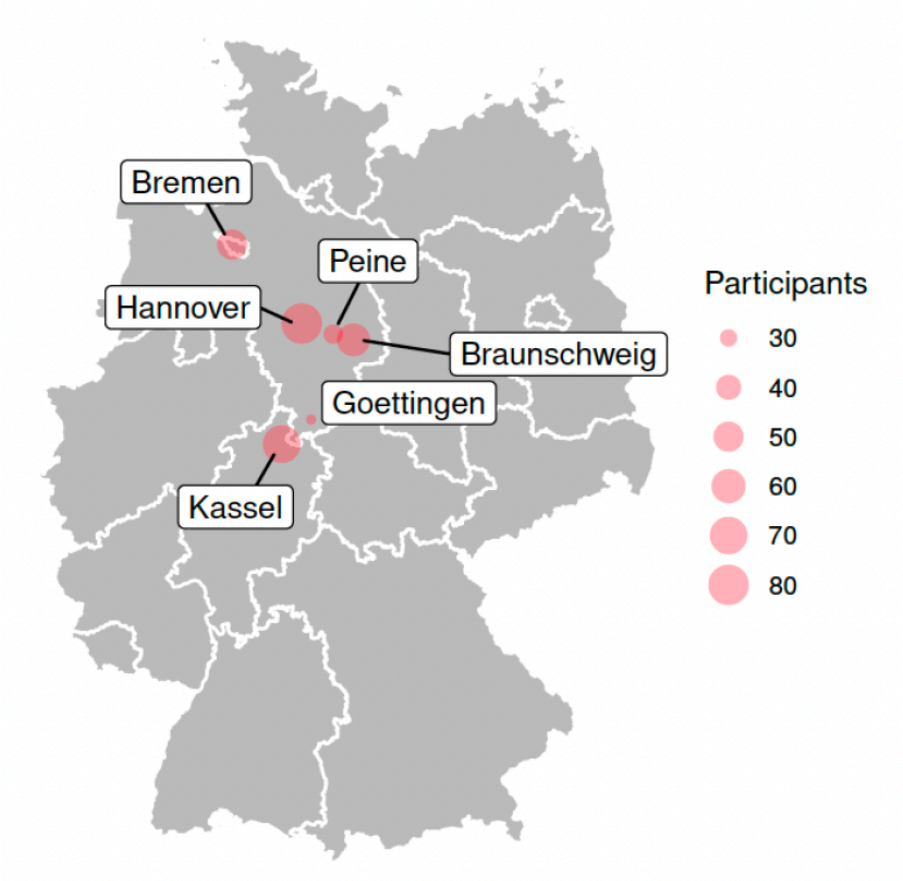

Figure 3: Map of Participating Cities

It is important to note that this analysis relies on a fraction of the originally planned sample size. The main study was scheduled for 2019-2020 but had to be terminated due to the COVID-19 pandemic and associated restrictions on in-person education. This constraint affects the statistical power of the analysis and precludes the planned heterogeneous treatment effect analyses by sex, migrant versus refugee status, education level, and age group.

5.2 Outcome Variables

The study examines four categories of outcomes. First, intermediate treatment outcomes measure take-up of the intervention through MALL usage—including any app use within the past two months (nine brackets of usage intensity), specific use in the past seven days, and whether particular recommended apps were downloaded.

Second, German language proficiency is measured through in-class tests administered at the end of CEFR levels A1 and A2, teachers' personal assessments of student progress, and (in ongoing data collection) long-term pass rates on the official integration course completion examination.

Third, self-reported German and communication skills are captured using Likert scales covering overall German skills, vocabulary, listening comprehension, reading, grammar, and ability to express oneself. An index of confidence in speaking German in public is also constructed.

Fourth, self-reported social integration is measured using an index based on the BiB Integration Survey, including time spent with people of German background, self-reported integration and satisfaction levels, lifestyle and political orientation, and future plans.

5.3 Balance Tests

Table 1 presents summary statistics by treatment status. The control group (N=170) and treatment group (N=147) are broadly comparable across observable characteristics. Mean age is approximately 32 years in both groups, and about 39% have refugee status in Germany. Average time in Germany is approximately 28 months, and about 70% have family members in Germany. Nearly all participants (97%) own a smartphone.

Some statistically significant differences emerge between groups. Treatment participants are more likely to speak German at home (10.9% vs. 18.5%, $p<0.10$) and less likely to have high education (19.4% vs. 11.6%, $p<0.10$) or vocational education (12.9% vs. 21.8%, $p<0.05$). There is also imbalance in the share of A2-level students (64.7% control vs. 51.0% treatment, $p<0.05$). The joint F-test for all characteristics yields an F-statistic of 1.406, indicating no systematic imbalance at conventional significance levels. Nevertheless, the empirical specifications include controls for these baseline characteristics to improve precision and address potential concerns.

Table 1: Summary Statistics by Treatment Status

Variable	(1) control N	(1) control Mean/SD	(2) treatment N	(2) treatment Mean/SD	Difference (1)-(2)	Normalized difference (1)-(2)
gender	168	0.363 (0.495)	146	0.329 (0.486)	0.034	0.070
age	163	32.350 (9.777)	138	32.913 (9.973)	-0.563	-0.057
refugee status in Germany	163	0.380 (0.487)	140	0.393 (0.490)	-0.012	-0.026
time in Germany (month)	150	28.293 (43.602)	146	27.596 (37.574)	0.697	0.017
family in Germany	165	0.697 (0.461)	143	0.699 (0.460)	-0.002	-0.005
mostly speak German at home	162	0.185 (0.390)	138	0.109 (0.312)	0.076*	0.214
no_educ	170	0.065 (0.247)	147	0.041 (0.199)	0.024	0.106
primary_educ	170	0.041 (0.199)	147	0.014 (0.116)	0.028	0.166
middle_educ	170	0.094 (0.293)	147	0.170 (0.377)	-0.076**	-0.226
high_educ	170	0.194 (0.397)	147	0.116 (0.321)	0.078*	0.215
vocational_educ	170	0.129 (0.337)	147	0.218 (0.414)	-0.088**	-0.234
bachelor_educ	170	0.306 (0.462)	147	0.272 (0.447)	0.034	0.074
master_educ	170	0.153 (0.361)	147	0.129 (0.337)	0.024	0.068
own smartphone	166	0.970 (0.171)	143	0.979 (0.144)	-0.009	-0.057
have to pay fees for integration course	164	0.433 (0.497)	144	0.389 (0.489)	0.044	0.089
share A2-level	170	0.647 (0.479)	147	0.510 (0.502)	0.137**	0.277
F-test of joint significance (F-stat)					1.406	
F-test, number of observations					250	

Notes: The table reports means with standard deviations in parentheses for the control group (column 1) and treatment group (column 2). Column 3 shows the difference in means, and column 4 shows the normalized difference. The F-test of joint significance tests whether all characteristics jointly predict treatment status. *** p<0.01, ** p<0.05, * p<0.10.

6. Empirical Methodology

6.1 Intention-to-Treat Estimation

The primary analysis estimates intention-to-treat (ITT) effects using ordinary least squares regression. The estimating equation is:

$$Yisc = \beta_1 + \beta_2 Tisc + \beta_3 Xi + \varepsilon isc$$

where Yisc is the outcome variable of interest for individual i in school s and class c; Tisc is a binary indicator for treatment assignment; Xi is a vector of individual characteristics including whether German is spoken at home, education level, and A1/A2 course level; and εisc is the error term. Standard errors are clustered at the class level to account for within-class correlation induced by the cluster-randomized design.

The coefficient β_2 identifies the causal effect of being assigned to treatment, which captures the intent to treat regardless of actual compliance with the intervention. For binary outcomes such as app downloads, probit and logit models are also estimated for comparison.

6.2 Complier Average Causal Effect Estimation

To estimate the effect of actual MALL usage on outcomes, I employ instrumental variables estimation to recover the complier average causal effect (CACE), also known as the local average treatment effect (LATE). Treatment assignment serves as an instrument for MALL usage, which satisfies the relevance condition given the strong first-stage effects documented below.

The validity of this approach requires the exclusion restriction—that treatment assignment affects outcomes only through its effect on MALL usage. This assumption is plausible given that the intervention specifically encouraged app usage without providing other resources or information that might independently affect language learning or integration. The monotonicity assumption requires that treatment assignment weakly increases MALL usage for all individuals, which is reasonable given that the control group received no encouragement.

Given the multiple possible definitions of compliance in an encouragement design, I estimate CACE using three alternative compliance measures: (1) downloading the Ankommen app; (2)

using MALL on a daily basis; and (3) downloading at least three of the recommended apps. These different definitions capture varying intensities of engagement with the intervention.

7. Results

7.1 First-Stage Effects: App Adoption

Table 2 presents results for intermediate treatment outcomes measuring MALL adoption. App usage is high among all study participants, reflecting the widespread familiarity with smartphones in this population. However, the intervention produced substantial differential effects on adoption of the recommended applications.

Columns (3) and (4) show that treatment assignment increased the probability of downloading the Ankommen app by 34.9 to 36.4 percentage points, depending on the specification. Given that only 5% of control group participants downloaded this app, the treatment effect represents a ninefold increase in adoption. This large effect is consistent with the intervention successfully directing participants toward the specifically recommended integration-focused application.

Columns (5) and (6) examine the number of apps downloaded excluding the Ankommen app from the video. Treatment assignment increased the number of additional apps downloaded by approximately 1.0 to 1.2 apps, indicating that the intervention encouraged broader engagement with language-learning applications beyond the focal app.

The time use data provide additional evidence of behavioral change. On average, treatment group participants reported spending 30 minutes per day using language-learning apps, compared to 15 minutes per day in the control group—a doubling of learning time. These first-stage results demonstrate that the encouragement intervention successfully increased both the extensive and intensive margins of MALL usage.

Table 2: Intermediate Treatment Outcomes

	any app Probit		Ankommen Probit		# of apps from video (except Ankommen) OLS		monthly usage Logit	
	(1) No Controls	(2) Controls	(3) No Controls	(4) Controls	(5) No Controls	(6) Controls	(7) No Controls	(8) Controls
treatment	0.0345	0.00406	0.364***	0.349***	1.150***	1.031***	0.554*	0.349
	(0.0456)	(0.0474)	(0.0453)	(0.0519)	(0.230)	(0.260)	(0.326)	(0.333)
N	317	293	317	293	317	293	317	293
R^2					0.190	0.188		

Notes: Columns (1)-(2) report marginal effects from probit estimation for any app usage. Columns (3)-(4) report marginal effects from probit estimation for downloading the Ankommen app. Columns (5)-(6) report OLS estimates for the number of apps downloaded excluding Ankommen. Columns (7)-(8) report marginal effects from logit estimation for monthly app usage. Controls include German spoken at home, education level, and A1/A2 course level. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

7.2 Intention-to-Treat Effects on Final Outcomes

Table 3 presents intention-to-treat effects on final outcomes. The self-reported German skills index (columns 1 and 2) contains standardized ratings for overall German skills, reading, listening, and vocabulary. Being in the treatment group increases this standardized index by 0.30 standard deviations ($p<0.05$ with controls). On the original scale from 4 to 20, treatment students reported an average score of 13.48, compared to 12.52 for control students.

The magnitude of this treatment effect is consistent with the existing literature. For the power calculations underlying this study, an effect size of 0.2 standard deviations was assumed based on prior research; the observed effect exceeds this benchmark. The effect is also economically meaningful, representing nearly one point improvement on the 16-point scale.

Columns (3) and (4) examine daily-life comfort in situations requiring German language use. The treatment coefficient is positive (0.17-0.19 standard deviations) but does not achieve statistical significance at conventional levels, likely reflecting limited statistical power.

Columns (5) and (6) show no significant effect on activities conducted in German. Similarly, columns (7) and (8) show no effect on the integration outcomes index. Several explanations may account for the null findings on integration measures: the effect may be too small to detect with the available sample size; the two-month treatment period may be too short for effects on broader integration to materialize; or the study may be underpowered to detect effects on these secondary outcomes. Long-term official state examination results are still being collected and may reveal effects that emerge over longer horizons.

Table 3: Final Treatment Outcomes—ITT Estimates

	German skills		Daily-life comfort		Activities in German		Integration outcomes	
	(1) No Controls	(2) Controls	(3) No Controls	(4) Controls	(5) No Controls	(6) Controls	(7) No Controls	(8) Controls
treatment	0.301**	0.298*	0.166	0.188	0.0415	0.0313	-0.116	-0.100
	(0.045)	(0.053)	(0.242)	(0.202)	(0.721)	(0.798)	(0.384)	(0.410)
N	312	288	302	279	304	281	307	283
R^2	0.023	0.050	0.007	0.070	0.000	0.078	0.003	0.078

Notes: All outcome variables are standardized indices with mean zero and standard deviation one. German Skills is an index of self-reported overall German skills, reading, listening, and vocabulary. Daily-Life Comfort measures comfort in daily situations requiring German. Activities in German captures the frequency of activities conducted in German. Integration Outcomes is an index based on the BiB Integration Survey including social contacts, satisfaction, and future plans. Controls include German spoken at home, education level, and A1/A2 course level. Standard errors clustered at the class level in parentheses. p-values shown in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

7.3 Complier Average Causal Effects

Table 4 presents CACE estimates for self-reported German skills using the three alternative compliance definitions. When compliance is defined as downloading the Ankommen app (columns 1 and 2), the estimated effect of actual MALL usage is 0.76 to 0.78 standard deviations ($p<0.10$). When compliance is defined as downloading at least three apps (columns 5 and 6), the effect is 0.63 to 0.83 standard deviations ($p<0.05$ with controls).

When compliance is defined as having used MALL on a daily basis (columns 3 and 4), no significant effect emerges. This null finding likely reflects the overall high rate of daily app usage in both treatment and control groups, which compresses the variation available for identifying effects among compliers.

These CACE estimates are substantially larger than the ITT effects, which is expected when compliance is partial. The estimated effects of 0.78 to 0.83 standard deviations translate to approximately 2.5 to 2.6 points on the 4-to-20 scale, representing a substantial improvement in self-perceived language proficiency among those who actually adopted the recommended apps.

Table 4: Self-Reported German Skills—CACE Estimates

	Downloaded Ankommen		Used MALL on daily basis		Downloaded at least 3 apps	
	(1) No Controls	(2) Controls	(3) No Controls	(4) Controls	(5) No Controls	(6) Controls
Ankommen	0.756*	0.782*				
	(0.055)	(0.088)				
app_daily			1.614	2.396		
			(0.144)	(0.219)		
compliance_number_app					0.633**	0.831**
					(0.031)	(0.019)
N	312	288	312	288	312	288
R^2	.	.	.	.	0.017	0.045

Notes: The table reports instrumental variables estimates of the effect of MALL usage on self-reported German skills, using treatment assignment as an instrument. Columns (1)-(2) define compliance as downloading the Ankommen app. Columns (3)-(4) define compliance as using MALL on a daily basis. Columns (5)-(6) define compliance as downloading at least three recommended apps. Controls include German spoken at home, education level, and A1/A2 course level. Standard errors clustered at the class level. p-values shown in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

8. Conclusion

This paper provides experimental evidence on the effectiveness of Mobile-Assisted Language Learning for improving language acquisition among migrants and refugees in Germany. Using a randomized controlled trial in integration courses, I find that a low-cost encouragement intervention—consisting of a motivational video and informational materials about language-learning apps—successfully increases app adoption and usage, and improves self-reported German language proficiency.

The results demonstrate substantial behavioral responses to the intervention. Treatment assignment increased adoption of the recommended Ankommen app from 5% to 46% and doubled daily time spent on language-learning activities from 15 to 30 minutes. These behavioral changes translated into improved language outcomes: intention-to-treat estimates show a 0.3 standard deviation improvement in self-reported German skills, while complier average causal effect estimates suggest that actual app adoption improves skills by 0.78 to 0.83 standard deviations.

These findings have important policy implications. The intervention tested here is highly scalable and low-cost, requiring only the distribution of informational materials and a brief motivational presentation. The reliance on freely available language-learning applications means there are no ongoing per-participant costs. Given the large population of migrants participating in integration

courses and the modest success rates on final examinations, even small improvements in language acquisition could yield substantial aggregate benefits.

Several limitations warrant acknowledgment. First, the COVID-19 pandemic terminated data collection before the target sample size was achieved, limiting statistical power and precluding planned heterogeneity analyses. Second, there is potential for spillover effects if flyers were shared between treatment and control participants, though class-level randomization should limit this concern. Third, the outcome measures for language proficiency are self-reported; long-term data on official examination results are still being collected and will provide more objective measures of learning gains. Fourth, the non-standardized nature of internal A1-level tests across classes introduces measurement noise. Finally, effects on broader integration outcomes were not detected, possibly due to limited power or the short time horizon of the study.

Future research should address these limitations through larger-scale implementations with longer follow-up periods to assess effects on objective language tests and labor market outcomes. The role of specific app features, optimal dosage of app usage, and heterogeneous effects across different migrant populations also merit investigation. Understanding how mobile learning can complement rather than substitute for traditional classroom instruction represents an important direction for optimizing integration course design.

In conclusion, this study demonstrates that mobile technology offers a promising, scalable tool for enhancing language acquisition among migrants. By lowering barriers to supplemental learning and providing flexible, accessible practice opportunities, language-learning applications can contribute to the broader goal of successful migrant integration.

References

Aldashev, A., Gernandt, J., & Thomsen, S. L. (2009). Language usage, participation, employment and earnings: Evidence for foreigners in West Germany with multiple sources of selection. *Labour Economics*, 16(3), 330-341.

BAMF. (2019). *Das Bundesamt in Zahlen 2018*. Bundesamt für Migration und Flüchtlinge.

Bleakley, H., & Chin, A. (2004). Language skills and earnings: Evidence from childhood immigrants. *Review of Economics and Statistics*, 86(2), 481-496.

Bleakley, H., & Chin, A. (2010). Age at arrival, English proficiency, and social assimilation among US immigrants. *American Economic Journal: Applied Economics*, 2(1), 165-192.

Burston, J. (2015). Twenty years of MALL project implementation: A meta-analysis of learning outcomes. *ReCALL*, 27(1), 4-20.

Chiswick, B. R., & Miller, P. W. (2015). International migration and the economics of language. In B. R. Chiswick & P. W. Miller (Eds.), *Handbook of the Economics of International Migration* (Vol. 1, pp. 211-269). Elsevier.

Cho, K., Lee, S., Joo, M. H., & Becker, B. J. (2018). The effects of using mobile devices on student achievement in language learning: A meta-analysis. *Education Sciences*, 8(3), 105.

Dörnyei, Z., & Chan, L. (2013). Motivation and vision: An analysis of future L2 self images, sensory styles, and imagery capacity across two target languages. *Language Learning*, 63(3), 437-462.

Dustmann, C., & Fabbri, F. (2003). Language proficiency and labour market performance of immigrants in the UK. *Economic Journal*, 113(489), 695-717.

Eldridge, S. M., Costelloe, C. E., Kahan, B. C., Lancaster, G. A., & Kerry, S. M. (2015). How big should the pilot study for my cluster randomised trial be? *Statistical Methods in Medical Research*, 25(3), 1039-1056.

Foged, M., Hasager, L., & Peri, G. (2022). Comparing the effects of policies for the labor market integration of refugees. *NBER Working Paper* No. 30534.

Foged, M., Hasager, L., Peri, G., Arendt, J. N., & Bolvig, I. (2023). Language training and refugees' integration. *Review of Economics and Statistics*, forthcoming.

Foged, M., & van der Werf, C. (2023). The integration of refugees: Evidence from Denmark. *Working Paper*.

German Council of Economic Experts. (2022). *Annual Report 2022/23*. Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung.

Hassan Taj, I., Sulan, N. B., Sipra, M. A., & Ahmad, W. (2016). Impact of mobile assisted language learning (MALL) on EFL: A meta-analysis. *Advances in Language and Literary Studies*, 7(2), 76-83.

Heller, S., & Slungaard Mumma, K. (2023). Learning to communicate: The effects of adult English language learning. *NBER Working Paper* No. 31702.

Kanas, A., & Kosyakova, Y. (2023). Language acquisition and the economic integration of refugees in Germany. *Journal of Ethnic and Migration Studies*, 49(11), 2880-2904.

Kosse, F., Deckers, T., Pinger, P., Schildberg-Hörisch, H., & Falk, A. (2020). The formation of prosociality: Causal evidence on the role of social environment. *Journal of Political Economy*, 128(2), 434-467.

Lochmann, A., Rapoport, H., & Speciale, B. (2019). The effect of language training on immigrants' economic integration: Empirical evidence from France. *European Economic Review*, 113, 265-296.

Muir, C. (2018). *Motivation and mindset in the foreign language classroom*. Unpublished doctoral dissertation, University of Nottingham.

Muir, C., Dörnyei, Z., & Adolphs, S. (2021). Role models in language learning: Results of a large-scale international survey. *Applied Linguistics*, 42(1), 1-23.

Resnjanskij, S., Ruhose, J., Wiederhold, S., Woessmann, L., & Wedel, K. (2024). Can mentoring alleviate family disadvantage in adolescence? A field experiment to improve labor-market prospects. *Journal of Political Economy*, forthcoming.

Ruhose, J. (2013a). Education, language, and the integration of immigrants in Germany. *Working Paper*.

Sayman, S., & Öncüler, A. (2008). An investigation of time-inconsistency. *Management Science*, 55(3), 470-482.

Schmid, S. D. (2023). Language as a barrier? Immigration and language learning in Germany. *European Sociological Review*, forthcoming.

Schuss, E. (2018). The impact of language course attendance on the employment of refugees in Germany. *IAB Discussion Paper* No. 35/2018.

Sung, Y. T., Chang, K. E., & Liu, T. C. (2015). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. *Computers & Education*, 94, 252-275.

Thompson, A. S., & Vásquez, C. (2015). Exploring motivational profiles through language learning narratives. *Modern Language Journal*, 99(1), 158-174.

Williamson, I., Leeming, D., Lyttle, S., & Johnson, S. (2013). Evaluating the audio-diary method in qualitative research. *Qualitative Research Journal*, 15(1), 20-34.

Yao, Y., & Van Ours, J. C. (2015). Language skills and labor market performance of immigrants in the Netherlands. *Labour Economics*, 34, 76-85.