Does study language (Dutch versus English) influence study success of Dutch and German students in the Netherlands?

Johanna F. de Vos,[1,2] Herbert Schriefers,[1] and
Kristin Lemhöfer[1]
[1] Radboud Universiteit | [2] International Max Planck Research School for
Language Sciences

We investigated whether the language of instruction (Dutch or English) influenced the study success of 614 Dutch and German first-year psychology students in the Netherlands. The Dutch students who were instructed in Dutch studied in their native language (L1), the other students in a second language (L2). In addition, only the Dutch students studied in their home country. Both these variables could potentially influence study success, operationalised as the number of European Credits (ECs) the students obtained, their grades, and drop-out rates. The L1 group outperformed the three L2 groups with respect to grades, but there were no significant differences in ECs and drop-out rates (although descriptively, the L1 group still performed best). In conclusion, this study shows an advantage of studying in the L1 when it comes to grades, and thereby contributes to the current debate in the Dutch media regarding the desirability of offering degrees taught in English.

Keywords: English-medium education, Dutch-medium education, higher education, study language, second language acquisition, study success, academic achievement, content learning

1. Introduction

1.1 Studying in a second language

In the Dutch media it is regularly argued that lectures and classroom interactions are not of the same quality when they are held in English as compared to Dutch (e.g., Hermans, 2017; Huygen, 2017; Kleinjan, 2017). This may influence how

much students can learn from lectures and classroom interactions. In addition, De Groot (2017, p.14) argues that the higher mental load that second language (L2) users experience is likely to negatively impact information processing and knowledge transfer. In this article, which is a shortened version of a chapter in De Vos's (2019) PhD thesis, we therefore address the question of whether studying in an L2 leads to worse study outcomes than studying in one's native language (L1).

Outside of the Dutch context, similar questions were previously evaluated by Macaro, Curle, Pun, An and Dearden (2018) in a systematic review, focusing on the use of English as the medium of instruction (EMI) in higher education. The evidence they found was mixed. Sometimes the students' listening comprehension was lower in L2 English than their L1 (e.g., Hellekjaer, 2010), but sometimes no difference could be detected (e.g., Joe & Lee, 2013). With regard to the students' grades, both Dafouz, Camacho and Urquia (2014), and Tatzl and Messnarz (2013) found no difference between groups of students who studied in L2 English or in their L1. However, for many of the studies included in their review, Macaro et al. (2018) question whether the conditions in the included studies were truly comparable. This led Macaro et al. to conclude that the evidence on the relationship between EMI and content learning is inconclusive.

In the Dutch context, Vinke (1995) investigated Dutch lecturers' and students' perceptions of using Dutch versus English as the language of instruction. In a survey held with 131 university lecturers, Vinke found that about 60% of them found the experience of teaching in both languages to be roughly comparable (Vinke, 1995, p.76). Still, the majority of lecturers also said they needed (much) more time to prepare courses in English, and felt less capable to express themselves, or to express something in a different way. About half of the respondents also felt less able to improvise in English than in Dutch. Furthermore, about half of the respondents assessed their own teaching quality to be lower in English (Vinke, Snippe & Jochems, 1998).

Vinke (1995) also videotaped 16 Dutch lecturers when they were teaching in Dutch and English, and had them be rated by observers. Four out of seven teaching behaviours under investigation were rated less favourably by the observers when the teaching was in English. These behaviours were body movement, variation in intonation and speed of delivery, verbal fluency, and the use of vague terms. No difference was found in the consultation of notes, the use of visual support, and the use of gestures.

Moving from Dutch lecturers to Dutch students, another question is whether studying in L2 English affects learning outcomes. To investigate this, Vinke (1995) let Dutch students watch a video-taped lecture either in Dutch or in English. The two lectures were given by the same Dutch lecturer, and contained the same content. Afterwards, the students answered true/false questions in Dutch about the

content of the lecture they had just seen. Out of 30 questions, the 34 students who had seen the lecture in Dutch on average scored 22.3 points, and the 34 students who had seen the lecture in English on average scored 21.0 points. This difference amounted to a small-to-medium effect size of $d = 0.45$, and was significant. The students in both conditions had been matched in terms of general academic ability.

Thus, Vinke's (1995) last study showed a detrimental effect of studying in L2 English when it comes to content learning, although the effect was relatively small. This outcome contrasts with that of Vander Beken and Brysbaert (2018), who let native speakers of Dutch (university students) read texts in Dutch and English. After reading, the students answered true/false recognition questions in the same language they had read the text in. No significant effect was found.

1.2 Studying abroad

Not only the study language may influence study success. Rienties, Beausaert, Grohnert, Niemandsverdriet and Kommers (2012) found that western international students in the Netherlands outperformed Dutch students in terms of grades and European Credits (ECs). They explain this finding with the fact that western international students in general are one or two years older than Dutch students, and that studying abroad is a conscious choice for them. As a result of the latter, these students are claimed to be more motivated and have better learning attitudes (Tempelaar, Rienties, & Gijselaers, 2006).

1.3 The present study

In this study, we investigated the relationship between study outcomes and study language, as well as the influence of studying abroad versus in one's home country. We did so using data that were provided to us by the Psychology programme of Radboud University in Nijmegen, the Netherlands. Since the academic year 2016–2017, this programme is offered in two tracks that are identical except for the instruction language. There is a track fully taught in English ('the English track'), and a track taught in Dutch ('the Dutch track') except for some study materials (articles and text books), which are in English. Because the curriculum and the content is exactly the same in the two tracks, this creates the perfect context to investigate the effects of study language on study success. 'Study success' was operationalised through the grades and the number of ECs the students obtained, as well as their drop-out rates.

We contrasted the two study languages (Dutch and English), and also two nationalities (Dutch and German). Thus, there were four groups of students in

the analysis: Dutch students in the Dutch track (studying in their L1, in their home country), Dutch students in the English track (studying in an L2, in their home country), German students in the Dutch track (studying in an L2, abroad), and German students in the English track (studying in an L2, abroad). Because information about the students' native language was not available, we worked from the assumption that the Dutch students spoke Dutch as their L1 and the German students German, though this assumption will not necessarily be correct for all the students in our sample.

The students were free to choose the language in which they wanted to study, and their nationality was a given. As a result, it is possible that there were pre-existing differences between the four groups which may have influenced study success outcomes. For example, it is conceivable that more intelligent and/or more highly motivated Dutch students are more likely to choose the L2 English track rather than the L1 Dutch track, as the former might be considered more challenging or prestigious. If this is true, it can impact our analysis of study success. For this reason, we first investigated whether there were any pre-existing differences between the Dutch students in the Dutch and English tracks by examining their mean high school exam grade, which is a known predictor of academic achievement in psychology students (e.g., De Koning, Loyens, Rikers, Smeets, & Van der Molen, 2012). To anticipate the results, we did not find such a difference between the two Dutch groups: they were comparable in terms of their high school performance.

In addition to a relationship between academic achievement and high school grades, a positive relationship between academic achievement and proficiency in the study language has been established (e.g., De Koning et al., 2012; Fonteyne, De Fruyt, & Duyck, 2014; Zijlmans, Neijt, & Van Hout, 2016). Therefore, we also compared the Dutch students in both tracks on their high school exam grade for English, since even in the Dutch track many of the reading materials were in English. As will be detailed below, we found that the Dutch students in the English track had scored significantly better on their English high school exam than the Dutch students in the Dutch track. Because such a pre-existing difference between the groups is undesirable, we matched the Dutch students in both tracks on their English high school grade. In this way, we ensured as much as possible that any effects of study language on study success would not be caused by underlying differences in English proficiency between the two groups.

After this correction for pre-existing differences between the two Dutch groups, we investigated whether the four groups differed on three measures of study success (ECs, grades and drop-out rates). We expected to find superior outcomes for those students who studied in their L1 (i.e., the Dutch students in the Dutch track) as compared to the other three groups.

In summary, we addressed the following research questions:

1. Are the 'better' Dutch students (i.e., those with higher high school exam grades) more likely to choose the L2 (English) rather than the L1 (Dutch) track?
2. Is there a relation between nationality, study language and study success?

2. Methods

2.1 Participants

Our participants were 614 first-year psychology students at Radboud University, who either had the Dutch or German nationality (but not both). These students had all received an e-mail explaining the current study, and had not objected to the use of their data. This approach to obtaining consent had been approved by the ethical committee of the Social Science faculty of Radboud University. We excluded 31 students who had deviated from the standard first-year psychology curriculum, for example who had taken extra courses. The demographic information of the remaining 583 students is shown in Table 1.

Table 1. Demographic information of the 583 students

Nationality	Track	n	% female	Mean age (SD)	Age range
Dutch	Dutch	172	80%	19.52 (2.64)	17–47
	English	36	67%	19.17 (1.20)	17–23
German	Dutch	50	76%	20.86 (1.88)	18–26
	English	325	68%	20.65 (1.92)	17–30
	Total	**583**	**72%**	**20.24 (2.20)**	**17–47**

2.2 Data

The data set consisted of demographic information (date of birth, study language, nationality), and for some of the Dutch students their high school exam grades. The data set also contained the grades and the number of ECs that the students had obtained in the 13 courses that make up the programme of the first year of Psychology at Radboud University. Grades are given on a 1–10 scale, where a 6 or higher is needed to pass. In case a student passes a course, he/she obtains all the ECs which that course is worth (one EC represents a work load of 28 hours). Otherwise, he/she obtains no ECs at all. We also had information about whether

or not a student dropped out of the Psychology programme before entering the second year. The data set and analysis script can be found at https://github.com /johannadevos/StudyLanguage.

2.3   Analysis

2.3.1   *Investigating pre-existing group differences*

Question 1 asks whether 'better' Dutch students (i.e., those with better high school exam grades) are more likely to opt for the English track. To answer this question, we compared the Dutch students in the two tracks on two dependent variables: their mean high school exam grade over all courses, and their high school exam grade for English. The independent variable was track (two levels: Dutch, English). Note that the dependent variables were measured *before* the students enrolled in either the Dutch or English track. Therefore, if we find significant pre-existing differences between the students in the two tracks, we can infer that track choice is related to general (high school) academic ability, and/or English proficiency. This question was restricted to the Dutch students because we could not obtain the high school data of the German students.

Since the high school grades mostly seemed to be non-normally distributed, we used the non-parametric Wilcoxon rank-sum test with continuity correction to compare means. We present bootstrapped descriptive statistics (standard deviations and confidence intervals) so that there is no need to rely on distributional assumptions (also see the next section).

2.3.2   *Predicting ECs and grades*

Question 2 asks whether the four groups differed on the three measures of study success (ECs, grades and drop-out rates). The four groups were Dutch students in the Dutch track, Dutch students in the English track, German students in the Dutch track, and German students in the English track. We chose to directly compare study success between the four groups, as opposed to the option of having a two-by-two design with nationality (Dutch versus German) and track (Dutch versus English) as the independent variables. We preferred this design because the sample sizes of the four groups were quite different (see Table 1). This means that any potential main effects of nationality and track could not be meaningfully interpreted.

We explored the use of linear mixed-effects models and analysis of variance (ANOVA) to compare the four groups on ECs. This is detailed in De Vos (2019) and summarised here. All statistical analyses were carried out in R (R Core Team, 2018). The model diagnostics of the linear mixed-effects models showed that this

technique was not suitable for modelling the number of ECs. The residual plot showed a positive relationship between the predicted values and the residuals, while no such patterns should be visible in a good model. In addition, more than 5% of the model's predictions fell outside of the theoretical 95% error bounds. Using ANOVA to model ECs was not an acceptable option either, because this variable was clearly non-normally distributed in all four groups (the most common score was the maximum score of 60 ECs).

A solution for modelling data sets that cannot be approximated well by parametric models is to use bootstrapping. The basic idea behind this approach is that rather than making assumptions about the shape of the distribution of the data that one is modelling, this distribution is actually estimated from the data. This is done by taking a large number of samples of size N (with replacement) from the original study's sample, where N equals the original study's sample size (Efron & Tibshirani, 1993). Field and Wilcox (2017) stress that the use of such models is much preferred over the use of linear models whose assumptions are not met. Therefore, we performed an ANOVA with bootstrapping on the number of ECs. Such an ANOVA is called a *robust ANOVA*, because its test statistics are not dependent on the data being normally distributed. We also used bootstrapping to obtain standard deviations and confidence intervals for descriptive statistics. Whenever we used bootstrapping, we used 10,000 iterations.

Grades could be modelled satisfactorily with a linear mixed-effects model and with a 'regular', parametric ANOVA (see De Vos, 2019). However, in order to preserve consistency between the analyses, we decided to also use robust ANOVA for modelling grades.

2.3.3   *Predicting drop-out rates*

To investigate the effect of group on drop-out rates, we ran a logistic regression (because drop-out is a binary measure). After running this model, we checked the model diagnostics, following Field, Miles and Field (2012, p. 341). They are reported in De Vos (2019, pp. 201–202), and there was no reason for concern: Most importantly, there were no outliers, and the values of DFBeta for all observations and parameters were far away from the critical value of 1 (Field et al., 2012, p. 340), the largest one being 0.13. DFBeta is a measure of influence, reflecting how model parameters change when one data point is excluded from the analysis (Field et al., 2012, p. 270).

2.3.4   *Controlling Type-I error rates*

We adopted a very strict criterion for controlling Type-I error rates, the motivation being that it is important to avoid false positives in case policy makers would potentially base decisions for educational policies on the outcomes of the present

study or related studies. Controlling Type-I error rates was done in two consecutive steps.

First, we corrected α (.05) for the use of three dependent variables in the same research questions, which increases the chance of finding a chance result (see Lakens, 2016). To this end, we used Nyholt's (2004) software which calculates the significance threshold for a set of dependent variables, where the correlation between these variables is taken into account. This method ensures that Type-I error rates stay at 5%, while it is less conservative than a Bonferroni correction. For example, since Question 1 was evaluated on two dependent variables, a Bonferroni-corrected α would be .0250. However, Nyholt's (2004) significance threshold was .0283 and Li and Ji's (2005) threshold was .0253. Nyholt (2015) states that Li and Ji's threshold is "reportedly more accurate", although it should only be used if it is below Nyholt's (2004) threshold. Thus, in this case we set α to .0253. In the Results section, we present the new α level along with each research question.

In a second step, we also corrected α whenever we made multiple comparisons as a follow-up to the robust ANOVAs. This was done with the Benjamini & Hochberg (1995) procedure, which focuses on controlling the false discovery rate, and again is less conservative than the Bonferroni procedure. The *p*-values from all the contrasts are first ordered from smallest to largest. The rank of a *p*-value is *j*, where the smallest *p*-value has rank 1 and the largest has rank *k*. The corrected level of α, per *p*-value, is $j/k * α$.

3.   Results

3.1   Question 1: Do the better Dutch students opt for the L2 track?

Question 1 was evaluated on two dependent variables: the students' average high school exam grade over all courses, and their average high school English exam grade. These grades are shown in Table 2, for the Dutch students in both tracks (i.e., in the Dutch track and the English track). Nyholt's (2004) significance threshold was .0283 and Li and Ji's (2005) threshold was .0253. Thus, we set α to .0253, and consequently used 97.47% confidence intervals.

The Dutch students who had opted for the English track had obtained a significantly higher English grade in high school ($W = 4692$, $p < .001$). Hedges' *g* was 0.75, indicating a medium-sized effect. There was no significant difference in overall high school grades between the students in both tracks ($W = 4024.5$, $p = .07$), although the trend again was in favour of the Dutch students in the English track, with a (very) small effect size of $g = 0.19$.

Table 2. High school exam grades of the Dutch students

High school exam	Track	n	Mean (SD)	97.47% CI	Median	Range
English	Dutch	174	6.87 (0.84)	6.72–7.01	7	5–9
	English	39	7.49 (0.75)	7.21–7.74	7	6–9
Overall (all courses)	Dutch	174	6.69 (0.49)	6.61–6.78	6.58	5.89–8.50
	English	39	6.78 (0.40)	6.65–6.94	6.73	6.08–7.85

Note. The sample sizes are slightly different from those in Table 1, because students who deviated from the standard psychology curriculum had not been excluded for this research question. High school grades were available for the majority, but not all of the Dutch students.

3.2   Question 2: Do study language and nationality affect study success?

3.2.1   *Data set*

In answering Question 1, we had found that the two Dutch groups were unequal from the beginning in terms of their pre-existing knowledge of English, which would be problematic for the rest of the analyses. Contrary to what is often thought, such pre-existing differences cannot simply be controlled for by using analysis of covariance (ANCOVA), because independence of the covariate (here: high school English grade) and treatment (here: group) effect is a prerequisite of ANCOVA (Field et al., 2012, pp. 464–466). As Field et al. (2012) explain, pre-existing group differences should be resolved in one of two ways: either randomising participants to experimental groups (which was not an option for us), or matching the groups on the covariate.

Therefore, we conducted a procedure to match the two Dutch groups on their English grade. The details of this matching procedure are explained in De Vos (2019), but in summary, we first switched back to the data set that did not contain any students with deviating curricula. Then, we removed students from the larger group (Dutch students in the Dutch track) from the data set until the two Dutch groups had the same average English high school grade. In total, we removed 62 students from the data set, which left 94 Dutch students in the Dutch track for analysis, and 33 Dutch students in the English track. In the end, the matched data set contained 127 Dutch students and 375 German students, a total of 502 students.

The matching procedure had not targeted the students' overall exam grade, because we did not detect any pre-existing group differences on overall exam grade. In the matched data set the Dutch students in the two tracks still did not differ significantly on their overall exam grade: $M = 6.86$ ($SD = 0.48$) for the Dutch students in the Dutch track, and $M = 6.77$ ($SD = 0.39$) for the Dutch students in the English track ($W = 1429$, $p = .50$).

For the analysis of ECs and grades, we removed 71 students who had dropped out during the first year from this matched data set. This is because the students who dropped out took fewer courses, and therefore obtained fewer ECs. Their impending drop-out may also have impacted their grades. This could obscure the analysis of the effect of group. In the resulting data set of 431 students, the two Dutch groups still did not differ significantly on English high school grade ($W = 1381.5$, $p = .82$) or mean overall high school grade ($W = 1182$, $p = .31$).

3.2.2 Alpha

Question 2 was evaluated on three dependent variables. This time, Nyholt's (2004) threshold was .022 and Li and Ji's (2005) threshold was .0253. Since Li and Ji's estimate was not below Nyholt's, we set α to .022 for all analyses belonging to Question 2.

3.2.3 ECs

Table 3 contains the descriptive statistics for the number of ECs that was obtained in the four groups. The differences between the groups were non-significant according to the robust ANOVA with bootstrapping ($F(3, 427) = 3.24$, $p = .06$).

Table 3. Descriptive statistics for ECs

Group	n	Mean (SD)	97.8% CI	Median	97.8% CI	Range
Dutch in Dutch track	87	54.26 (9.93)	51.18–56.28	60	54–60	11–60
Dutch in English track	31	49.52 (15.71)	40.68–54.48	54	38–54	0–60
German in Dutch track	40	49.25 (13.71)	42.85–53.27	54	46.50–55.50	7–60
German in English track	273	50.29 (15.18)	47.98–52.22	60	54–60	0–60
Total	**431**	50.94 (14.40)	49.21–52.42	60	54–60	0–60

3.2.4 Grades

Descriptive statistics for grades are given in Table 4.

This time, the robust ANOVA showed that the differences between the groups were significant ($F(3, 424) = 4.62$, $p = .007$). The outcomes of robust post-hoc tests are shown in Table 5. After correcting α for multiple comparisons, these tests showed that the Dutch students in the Dutch track scored significantly better than

Table 4. Descriptive statistics for grades

Group	n	Mean (SD)	97.8% CI	Median	97.8% CI	Range
Dutch in Dutch track	87	7.23 (0.81)	7.02–7.43	7.25	6.96–7.44	5.18–9.02
Dutch in English track	31	6.68 (1.08)	6.16–7.07	6.84	6.44–7.09	3.77–8.60
German in Dutch track	40	6.73 (0.97)	6.35–7.07	6.85	6.34–7.19	3.88–8.79
German in English track	270	6.91 (1.08)	6.75–7.05	7.10	6.85–7.22	3.41–8.81
Total	**428**	**6.94 (1.02)**	**6.82–7.05**	**7.10**	**6.95–7.17**	**3.41–9.02**

all other groups. In contrast, the other three groups did not significantly differ from one another.

Table 5. Pairwise comparisons between all groups on Mean grade

Contrast	Mean difference (SE)	97.8% CI	p	Corrected α
Dutch in Dutch track – Dutch in English track	0.55 (0.22)	0.09–1.05	**.005**	3/6 * .022 = .011
Dutch in Dutch track – German in Dutch track	0.50 (0.18)	0.08–0.90	**.004**	2/6 * .022 = .007
Dutch in Dutch track – German in English track	0.32 (0.11)	0.07–0.57	**.0028**	1/6 * .022 = .004
Dutch in English track – German in Dutch track	−0.05 (0.25)	−0.63–0.49	.85	6/6 * .022 = .022
Dutch in English track – German in English track	−0.23 (0.21)	−0.72–0.22	.27	4/6 * .022 = .015
German in Dutch track – German in English track	−0.18 (0.17)	−0.57–0.20	.29	5/6 * .022 = .018

Note. Alpha levels were corrected with the Benjamini & Hochberg (1995) procedure. Significant *p*-values are printed in bold.

3.2.5 *Drop-out*

Table 6 shows the drop-out rates per group.

Table 6 suggests that the German students dropped out more often than Dutch students. To statistically test this observation, we ran a logistic regression with group as the predictor, and drop-out (yes/no) as the dependent variable. Table 7 shows all six contrasts, including the corrected level of α. None of the contrasts were significant at this conservative α level.

Table 6. Percentage of students in each group who dropped out before the second year

Group	n	Drop-out rate	98% CI
Dutch in Dutch track	94	19.15%	9.69–28.61
Dutch in English track	33	21.21%	4.63–37.79
German in Dutch track	50	34.00%	18.39–49.61
German in English track	325	33.23%	27.14–39.32
Total	**502**	**29.88%**	**25.12–34.64**

Table 7. Planned contrasts between all groups

Contrast	b	SE	z	p	Corrected α
Dutch in Dutch track – Dutch in English track	0.13	0.50	0.26	.80	5/6 * .022 = .018
Dutch in Dutch track – German in Dutch track	0.78	0.40	1.96	.05	2/6 * .022 = .007
Dutch in Dutch track – German in English track	0.74	0.29	2.58	.01	1/6 * .022 = .004
Dutch in English track – German in Dutch track	0.65	0.52	1.25	.21	4/6 * .022 = .015
Dutch in English track – German in English track	0.61	0.44	1.39	.16	3/6 * .022 = .011
German in Dutch track – German in English track	−0.03	0.32	0.11	.91	6/6 * .022 = .022

Note. Alpha levels were corrected for multiple comparisons with the Benjamini & Hochberg (1995) procedure.

4.   Discussion

4.1   Do the better Dutch students opt for the L2 track?

There was no significant difference in overall high school exam grades between the Dutch students who chose to study psychology in Dutch and in English. Thus, it seems that the general academic abilities (as far as they are reflected in high school grades) of the two groups were the same. On the other hand, the Dutch students who chose to study in English rather than Dutch had obtained a significantly higher English grade on their high school exam. This is not surprising, as Dutch students with relatively weak English skills would not prefer to study in a completely English track. Still, even in the Dutch track, many of the study materials (e.g., the text books) were in English. Therefore, the fact that we found a significant pre-existing difference in English skills between the two groups was relevant when comparing study success between those groups.

4.2   Do study language and nationality affect study success?

In our second research question, we investigated whether there were differences in study success between Dutch students in the Dutch track (i.e., students who studied in their L1), Dutch students in the English track (i.e., students who studied in an L2), and German students in the Dutch and English track (i.e., students who studied in an L2, and outside their home country). To be able to compare study success between the two groups of Dutch students without their English skills being a confounding variable, we matched the Dutch students in the two groups on their English exam grade. We did not have access to the high school grades of the German students, which means we cannot be equally sure about the comparability of the German students in the two tracks. In future studies, such information would obviously be desirable. For the current study, the comparisons involving the German groups nevertheless are interesting for university lecturers and policy makers.

4.2.1   *ECs*

We did not detect a significant effect of group on the number of obtained ECs, although descriptively the Dutch students in the Dutch track outperformed the three other groups (and the *p*-value was .06, at an α level of .022). Please recall that students earn ECs when obtaining a grade of 6 or higher. The mean grade obtained in all four groups was amply above this cut-off point, the lowest mean grade being 6.68 for the Dutch students in the English track. Thus, the number of ECs may have been too coarse a measure to distinguish between the students in the four tracks. The ECs variable remains relevant because it underlies policies such as the *Bindend Studieadvies* (literally in English: *Binding Study Advice*). This policy states that students are only allowed to continue into the second year of study when they have obtained a certain number of ECs in the first year (42 ECs for Psychology at Radboud University).

4.2.2   *Grades*

The Dutch students in the Dutch track obtained the highest grades, significantly outperforming all other groups. At least for the comparison between the Dutch students in the Dutch and in the English track, differences in high school grades cannot be responsible for this result, because we had created subgroups that were matched on high school grades. This outcome shows that there indeed seems to be an advantage of studying in the L1, even when a lot of the study materials are in English. This raises the question of what exactly caused the L1 advantage. It may be that the students' language skills enabled them to better study, remember and reproduce the material, but it is also possible that the advantage came from

listening to lecturers who were teaching in their L1 (in the Dutch track, most, if not all courses were taught by native Dutch lecturers). Our finding is in line with Vinke (1995), who found that Dutch students scored better on true/false content questions after watching a lecture in Dutch rather than English (the lecturer was a native speaker of Dutch). However, that study cannot disentangle the two potential explanations either.

Some of the other studies that we mentioned in the Introduction do provide evidence for one of the potential explanations. At the student level, Hellekjaer (2010) and Dafouz et al. (2014) found evidence that students' listening comprehension is lower in the L2. At the lecturer level, a study in Vinke's (1995) dissertation showed that native Dutch lecturers believe their teaching is of lower quality in English as compared to Dutch. Dutch lecturers were also rated less favourably by observers when they were teaching in English rather than Dutch.

Of course, the two explanations are not mutually exclusive. In the context of Dutch higher education, it would be interesting to determine their relative importance. This would yield more insight in the question of whether students' academic achievements benefit more from offering English language training to students and/or to lecturers. An interesting experiment could be to let Dutch students attend a lecture that is either given by an L1 Dutch speaker in Dutch or an L1 English speaker in English, and have them answer content questions afterwards. This would ensure that any effects would be due to the students' language proficiency, and not the nativeness of the lecturer. The comparability of the materials in the Dutch and English condition should be strictly controlled.

A different conclusion that one could draw from the same data is that the use of English in Dutch higher education should be restricted wherever possible. Irrespective of the outcomes of the current study, many people already argue for such a change, and for various reasons (e.g., Huygen, 2017; Teuling, 2017; Vasterman, 2017). Of course, such a decision would depend on many other factors as well, such as the wish to recruit non-Dutch speaking students or lecturers, or the desire to create an international classroom.

However, before implementing any new policies on the basis of the outcomes of this study, we should consider how much importance should be attached to the magnitude of the effects that we found, regardless of their significance. The largest mean grade difference between the four groups was 0.55 points on a 1–10 scale (between the Dutch students in the Dutch track and the Dutch students in the English track), corresponding to a medium effect size of $g = 0.62$. Such a difference may or may not warrant the costs of offering English training to students and/or lecturers, and may or may not warrant radical changes in language policy in Dutch higher education. In addition, longitudinal research is needed to investigate how the grade differences between the groups develop over the study period

of three years (i.e., the duration of a bachelor's degree), rather than the first year only.

4.2.3   *Drop-out*

The descriptive statistics showed quite substantial differences between the drop-out rates of Dutch and German students at Radboud University, but these differences did not reach significance. We should point out that our α level of .004 (corrected for multiple dependent variables, and multiple comparisons) was very strict, and as a result one *p*-value of .01 did not reach significance. Please recall that we opted for a very conservative α level to avoid basing high-impact policy decisions on potential false positives. Nevertheless, it seems important and worthwhile to continue monitoring drop-out rates between the different groups of students.

Another interesting matter is that the drop-out rates in the current study were quite high in an absolute sense (up to 34% for German students in the Dutch track). Therefore, it seems recommendable in general to provide more guidance to students when they are choosing their programme of study, and to continue to investigate what causes students to drop out. Of course, the reasons for students to drop out can be manifold. They will not always be related to students' choice of study programme, study language, or their academic achievements, and will therefore not always be preventable.

4.2.4   *No effect of studying abroad*

While we found an L1 advantage on grades, we found no significant difference between the three groups who studied in an L2 on any of the measures (although, as mentioned above, there was a trend in drop-out rates). In other words, the Dutch students in the English track did not differ from the German students in the Dutch or English track. This is not in line with the findings of Rienties et al. (2012), who found that western international students in the Netherlands obtained higher grades and more ECs than Dutch students (also see Section 1.2 in the Introduction).

Rienties et al. (2012) explain their findings by referring to the students' age (indeed, the German students in our sample also were slightly older than the Dutch students, see Table 1) and the idea that studying in the Netherlands is a more conscious choice for non-Dutch students. Perhaps in the current study these presumed advantages were cancelled out by the fact that German students who come to the Netherlands to study psychology often have not obtained a place to study in Germany because of the highly selective, (high-school) grade-based admission procedure in Germany. In other words, the academic ability of the German students' in our sample may have been slightly lower than that of the Dutch

students. To untangle such potential confounds, it is desirable that future investigations include data on the academic ability of all participants (in our data set, such information unfortunately was not available). Since only German students in a psychology programme were included in our study, we do not know whether the current findings also generalise to students of other nationalities, especially non-western students, and to other study programmes. This has also to be investigated in future research.

5.   Summary and conclusions

In the current study, we addressed two questions regarding the relationship between students' nationality, study language, and study success. To begin with, we showed that students who study in their L1 (here, Dutch students who study in Dutch) outperform students who study in an L2 (here, Dutch students who study in English, or German students who study in Dutch or English) when it comes to grades. In this study, the Dutch students in the Dutch track also obtained more ECs than the other three groups, but not significantly so. With regard to drop-out rates, the Dutch students (in both tracks) seemed to drop out less often than the German students (in both tracks), but again not significantly so. On none of the measures, there was a significant difference between the Dutch and German students in the English track (i.e., who studied in the same L2). This shows that, at least when comparing Dutch and German students, there was no effect of studying abroad.

All these results were obtained after the Dutch students in the Dutch and English tracks had been matched on their English high school grades – in our original sample, the Dutch students in the English track on average had obtained a better English grade in high school than the Dutch students in the Dutch track. The overall high school grade, however, did not differ between the two Dutch groups in the original sample.

Offering English-medium education in the Netherlands currently is the topic of fierce debate in Dutch newspapers and media. Among the many voices, hardly any come from empirical research. With the current study, we strove to add empirical, objective insights to the debate. Using data that were provided by the Psychology educational institute at Radboud University, we found evidence that studying in an L2 is disadvantageous when it comes to students' grades. For the Dutch students in our sample, we know that this effect was not caused by pre-existing differences in high school grades between those students who chose to study in Dutch versus English. Because the use of English in Dutch education is

advancing day by day, it is adamant that we extend our knowledge of the potential effects this may have on students' education, be these effects positive or negative.

Funding

References

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A new and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Dafouz, E., Camacho, M., & Urquia, E. (2014). 'Surely they can't do as well': A comparison of business students' academic performance in English-medium and Spanish-as-first-language-medium programmes. *Language and Education*, *28*(3), 223–236. https://doi.org/10.1080/09500782.2013.808661

De Groot, A. M. B. (2017). *Nederlands moet. Over meertaligheid en de verengelsing van het universitaire onderwijs* [transcript]. University of Amsterdam, Amsterdam. Retrieved from https://www.scienceguide.nl/wp-content/uploads/2017/11/Binnenwerk_150917.pdf

De Koning, B. B., Loyens, S. M. M., Rikers, R. M. J. P., Smeets, G., & Van der Molen, H. T. (2012). Generation Psy: Student characteristics and academic achievement in a three-year problem-based learning bachelor program. *Learning and Individual Differences*, *22*(3), 313–323. https://doi.org/10.1016/j.lindif.2012.01.003

De Vos, J. F. (2019). Naturalistic word learning in a second language. Unpublished doctoral dissertation. Radboud University, Nijmegen.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York, NY: Chapman & Hall. https://doi.org/10.1007/978-1-4899-4541-9

Field, A. P., & Wilcox, R. R. (2017). Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers. *Behaviour Research and Therapy*, *98*, 19–38. https://doi.org/10.1016/j.brat.2017.05.013

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. London, UK: SAGE.

Fonteyne, L., De Fruyt, F., & Duyck, W. (2014). To fail or not to fail? Identifying students at risk by predicting academic success. In M. F. Freda (Ed.), *Reflexivity in higher education: Research and models of intervention for underachieving students* (pp. 345–356). Rome, Italy: Aracne. https://doi.org/10.4399/978885487014727

Hellekjaer, G. O. (2010). Assessing lecture comprehension in Norwegian English-medium higher education. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 233–258). Amsterdam: John Benjamins. https://doi.org/10.1075/aals.7

Hermans, F. (2017, October 10). Studenten willen geen les meer in steenkolen-Engels van docenten. *De Gelderlander*. Retrieved from https://www.gelderlander.nl/nijmegen-e-o/studenten-willen-geen-les-meer-in-steenkolen-engels-van-docenten~a5a997b1/

Huygen, F. (2017, June 28). Opinie: Engels als voertaal vernielt het hoger onderwijs. *De Volkskrant*. Retrieved from https://www.volkskrant.nl/columns-opinie/opinie-engels-als-voertaal-vernielt-het-hoger-onderwijs~b7fd8359/

Joe, Y., & Lee, H.-K. (2013). Does English-medium instruction benefit students in EFL contexts? A case study of medical students in Korea. *The Asia-Pacific Education Researcher*, *22*(2), 201–207. https://doi.org/10.1007/s40299-012-0003-7

Kleinjan, G.-J. (2017, November 16). Steenkolenengels bij docenten: 'Experts die details en nuances missen. Zo jammer'. *Trouw*. Retrieved from https://www.trouw.nl/home/steenkolenengels-bij-docenten-experts-die-details-en-nuances-missen-zo-jammer-~ade0943b/

Lakens, D. (2016, February 14). *Why you don't need to adjust your α level for all tests you'll do in your lifetime* [Blog post]. Retrieved from http://daniellakens.blogspot.com/2016/02/why-you-dont-need-to-adjust-you-α.html

Li, J., & Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity*, *95*(3), 221–227. https://doi.org/10.1038/sj.hdy.6800717

Macaro, E., Curle, S., Pun, J., An, J., & Dearden, J. (2018). A systematic review of English medium instruction in higher education. *Language Teaching*, *51*(1), 36–76. https://doi.org/10.1017/S0261444817000350

Nyholt, D. R. (2004). A simple correction for multiple testing for SNPs in linkage disequilibrium with each other. *The American Journal of Human Genetics*, *74*(4), 765–769. https://doi.org/10.1086/383251

Nyholt, D. R. (2015). *Matrix Spectral Decomposition (matSpD) – Estimate the equivalent number of independent variables in a correlation (r) matrix*. Retrieved from https://sites.google.com/site/qutsgel/software/matspd-local-version

R Core Team (2018). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Rienties, B. C., Beausaert, S. A. J., Grohnert, T., Niemantsverdriet, S., & Kommers, P. (2012). Understanding academic performance of international students: The role of ethnicity, academic and social integration. *Higher Education*, *63*(6), 685–700. https://doi.org/10.1007/s10734-011-9468-1

Tatzl, D., & Messnarz, B. (2013). Testing foreign language impact on engineering students' scientific problem-solving performance. *European Journal of Engineering Education*, *38*(6), 620–630. https://doi.org/10.1080/03043797.2012.719001

Tempelaar, D. T., Rienties, B., & Gijselaers, W. H. (2006). Internationalisering: en de Nederlandse student? *Onderzoek van Onderwijs*, *35*(3), 40–45.

Teuling, I. (2017, July 11). *Zorgen over kwaliteit Engelstalig onderwijs universiteiten en hogescholen. De Volkskrant*. Retrieved from https://www.volkskrant.nl/nieuws-achtergrond/zorgen-over-kwaliteit-engelstalig-onderwijs-universiteiten-en-hogescholen~bcce8e6c/

Vander Beken, H., & Brysbaert, M. (2018). Studying texts in a second language: The importance of test type. *Bilingualism: Language and Cognition*, *21*(5), 1062–1074. https://doi.org/10.1017/S1366728917000189

Vasterman, P. (2017, November 13). Keuze voor Engels schaadt de inhoud op de universiteit. *NRC Handelsblad*. Retrieved from https://www.nrc.nl/nieuws/2017/11/13/keuze-voor-engels-schaadt-de-inhoud-op-de-universiteit-13989528-a1580911

Vinke, A.A. (1995). English as the medium of instruction in Dutch engineering education (Doctoral dissertation). Retrieved from Research Repository TU Delft (Accession no. uuid:491b55f9-fbf9-4650-a44d-acb9af8412a8)

Vinke, A.A., Snippe, J., & Jochems, W. (1998). English-medium content courses in non-English higher education: A study of lecturer experiences and teaching behaviours. *Teaching in Higher Education*, *3*(3), 383–394. https://doi.org/10.1080/1356215980030307

Zijlmans, L., Neijt, A., & Van Hout, R. (2016). The role of second language in higher education: A case study of German students at a Dutch university. *Language Learning in Higher Education*, *6*(2), 473–493. https://doi.org/10.1515/cercles-2016-0026

Address for correspondence

Johanna F. de Vos
johannadevos@gmail.com

Co-author information

Herbert Schriefers
Donders Centre for Cognition
Radboud Universiteit
h.schriefers@donders.ru.nl

Kristin Lemhöfer
Donders Centre for Cognition
Radboud Universiteit
k.lemhofer@donders.ru.nl