Validating GULS: An Open-Access Dutch Language Proficiency Test

Mona Bassleer[a], Stijn Schelfhout[a,b], Lot Fonteyne[c], Mit Leuridan[d], Wouter Duyck[a,e], Nicolas Dirix[a]

[a]Department of Experimental Psychology, Ghent University, 9000 Ghent, Belgium; [b]Department of Work, Organization and Society, Ghent University, 9000 Ghent, Belgium; [c]Department of Educational Affairs, Sint-Pietersnieuwstraat 33, Ghent University, 9000 Ghent, Belgium; [d]Teaching, Training and Doctoral Support, Ghent University, 9000 Ghent, Belgium; [e]The Accreditation Organization of the Netherlands and Flanders (NVAO), 2514 JG Den Haag, The Netherlands.

*Mona Bassleer

Address: Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium

Email: mona.bassleer@ugent.be

Abstract

Previous post-entry language assessments (PELAs) research in higher education shows that academic language proficiency contributes to academic achievement. PELAs are particularly valuable for higher education systems with minimal to no admission requirements to identify and support at-risk students. However, the availability of PELAs for languages like Dutch, beyond English, is limited. Moreover, existing Dutch PELAs and their construct validity evidence are not publicly accessible, and predictive validity analyses typically do not reach the program-specific level. Therefore, the present study introduced the Ghent University Language Screening (GULS), an easy-to-administer, free and publicly accessible Dutch PELA. More specifically, GULS evaluates reading comprehension of first-year students in higher education. First, we confirmed the construct validity of GULS at model and item level and its reliability using data from the two three-year periods 2017-2018 to 2019-2020 (N_1 = 12,527) and 2020-2021 to 2022-2023 (N_2 = 17,204). Second, we examined GULS's predictive validity for academic achievement (i.e., Grade Point Average and study success) on data over the same two periods across sixteen bachelor study programs (n_1 = 8,244; n_2 = 10,891), followed by predictive validity analyses for each study program across the combined six-year period. Results demonstrate that GULS is a valid and reliable PELA to assess Dutch language proficiency, especially for first-year higher education students requiring language support to ensure equal educational opportunities. As such, GULS functions as a predictor of first-year academic achievement. We discuss the possible application of GULS in future educational research and practice due to its accessibility.

Keywords: post-entry language assessment, academic achievement, higher education, validation

Validating GULS: An Open-Access Dutch Language Proficiency Test

Today's 21st-century society requires adequate language proficiency from individuals for a full-fledged societal participation (Kennedy & Sundberg, 2020; OECD, 2019). In addition to everyday language proficiency, academic language proficiency is important for students entering higher education (Knoch & Elder, 2013; Read, 2016). This addition aligns with Hulstijn's (2015) concept of Higher Language Cognition (HLC), which stands in contrast to Basic Language Cognition (BLC). BLC pertains to the language proficiency required for everyday communication, whereas HLC as equivalent of academic language proficiency (hereafter: language proficiency) involves more advanced language skills at lexical, syntactical, and cognitive levels (Hulstijn, 2015). Additionally, reading comprehension, as a component of language proficiency, appears to be a challenging receptive skill predominantly relied upon by students during the early stages of higher education (De Wachter et al., 2013; Jansen et al., 2022; Van Houtven et al., 2010). The present study considers language proficiency in terms of reading comprehension, defined as understanding, using, evaluating, and reflecting (on) textual content (OECD, 2023). Following Kintsch's (2013) Construction-Integration model, an individual constructs meaning from and integrates this textual content with existing long-term knowledge during the reading comprehension process. Furthermore, this ability relies on coordinating lower-order skills like decoding with higher-order ones such as vocabulary knowledge use (Kintsch, 2013).

Language proficiency is known for playing a crucial yet not exclusive role in academic achievement (Elder, 2017; Heeren et al., 2021; Read, 2016). This ability is indeed used across different subjects and academic disciplines for content comprehension (Read, 2015). Various meta-analytic evidence shows a clear association between language proficiency of mainly non-native speakers and academic achievement (i.e., Grade Point Average (GPA)), with average correlation coefficients ranging from $r = .18$ to $r = .23$

(Abunawas, 2014; Gagen, 2019; Ihlenfeldt & Rios, 2023; Wongtrirat, 2010). Nevertheless, language proficiency can challenge all (prospective) enrolling students in higher education (Elder, 2017; Wingate, 2015). In a recent meta-analysis, researchers specifically examined the reading comprehension components of more general admission assessments (e.g., SAT), showing a correlation coefficient of r = .29 (Clinton-Lisell et al., 2022). However, these instances of language proficiency assessment determine admission eligibility at English-medium higher education institutions.

Alternatively, higher education institutions with minimal or no admission requirements frequently employ post-entry language assessments (PELAs) to identify students at risk of academic underperformance. Indeed, the low-stakes admissions environments characteristic of these institutions tend to attract increasingly diverse student populations entering higher education (Elder, 2017; Knoch & Elder, 2013; Read, 2016). However, an unintended consequence of such open-access systems is the default unavailability of traditional indicators of academic preparedness, such as standardized test scores or high school GPA (see Meeter, 2023 and Westrick et al., 2015 for a full discussion on predictive validity of these indicators). As a result, institutions lack crucial information about the cognitive and academic skills of incoming students. In this context, PELAs can play a pivotal role in supporting student success by providing timely diagnostic insights and informing subsequent interventions. For the present study, the data on standardized exams is non-existent, while the data on high school GPA is unavailable.

To the best of our knowledge, research on the predictive validity of PELAs for all enrolled higher education students on academic achievement is rather scarce (e.g., De Wachter et al., 2013; Heeren et al., 2021; van Dijk, 2015). Moreover, the availability of PELAs for languages other than English is currently very limited. In general, a distinct relationship is observed in (Dutch) PELA studies, with language proficiency accounting for a maximum of 10% of the variance in academic achievement (Elder, 2017; Heeren et al., 2021; Knoch & Elder, 2013). The present study therefore focuses on the validation of a Dutch PELA. Dutch is the official language of higher education in Flanders and the Netherlands, with over 24,300,000 native speakers (Eberhard et al., 2024). Governing bodies explicitly advocate for maintaining the Dutch language proficiency in education and society, especially amidst the increasing language diversity (Jansen et al., 2022). Moreover, Dutch holds merits in scientific research exemplified by its status as one of the most studied languages in psycholinguistics (Siegelman et al., 2022).

Researchers highlight the practice of internally developed PELAs, yet often without professional validation (Knoch & Elder, 2013). A recent review similarly underscores that over half of their included language assessment studies fail to address construct validity, and those that do often limit their analysis to the overall model-data fit (Min & Aryadoust, 2021). The test developed by Heeren and colleagues (2021) is a good example of a Dutch PELA to screen first-year university students in an open-access[1] higher education system and identify those needing support. However, both the detailed construct validation evidence as well as the test itself are not publicly accessible, compromising transparency, replicability, and reproducibility (Min & Aryadoust, 2021). Additionally, the predictive validity of their PELA was examined on first-year academic achievement in terms of credit completion rate, whereas the most common metric of academic achievement in educational research is GPA (York et al., 2015). We include both GPA and study success (i.e., credit completion rate) as measures of academic achievement, recognizing the differing views on GPA's utility (Ihlenfeldt & Rios, 2023). Furthermore, assessing the predictive validity of language proficiency for

[1] Students who successfully completed secondary education can enroll in higher education without admission requirements like an entrance exam, except for the study programs Medicine, Dentistry, Veterinary Medicine, and Performing and Visual Arts.

academic achievement does not usually include separate analyses by study program. However, such methodological refinement can demonstrate the differential impact of language proficiency on academic achievement within distinct fields of study (Elder, 2017; Hauspie et al., 2024; Read, 2016).

To address the issues with and therefore advance existing PELA in higher education, the present study validates the Ghent University Language Screening (GULS), developed as a Dutch language proficiency PELA with fully open access (i.e., easy to administer, free and publicly available). More specifically, GULS assesses reading comprehension of first-year students in higher education. First, we examine and confirm GULS's construct validity on model and item level, and its reliability using extensive data from the academic years 2017-2018 to 2019-2020 and 2020-2021 to 2022-2023. Second, we evaluate the predictive validity of GULS for academic achievement (i.e., both GPA and study success) on large prospective data from the same two three-year periods across sixteen bachelor study programs, culminating in predictive validity analyses for each study program spanning this extended aggregate six-year period. As such, the present study provides insights into the contribution of Dutch language proficiency to first-year academic achievement in various study programs, such as linguistic but also social science, and STEM-oriented study programs. For these analyses, we also included control variables like socio-economic status (SES) and the mathematics category of the students' enrolled study program.[2] Practically, GULS can help inform (prospective) first-year higher education students about their language proficiency level, and guide remediation for those students at risk. Moreover, GULS's open accessibility supports its potential use in future educational research.

[2] The present study differentiates between study programs that expect basic or advanced levels of mathematical proficiency. The advanced mathematics study programs are thus linked to more advanced mathematical curricula (see also *Measures*).

Method

Participants

The present study used secondary data from first-year students at a large European university with eleven faculties and 42 bachelor programs, consistently ranked in the top 75 worldwide according to the Academic Ranking of World Universities (formerly Shanghai Ranking, see https://www.shanghairanking.com/rankings/arwu/2022). The FPPW Ethics Committee at Ghent University granted approval. The university's open-access system (see also Introduction) ensures a uniform curriculum for full-time first-year students within a study program. Our samples included students who entered higher education for the first time between 2017-2018 and 2022-2023 and participated in the university's study (re)orientation and remediation SIMON project. SIMON is a self-assessment tool that evaluates academic aptitude and vocational interests, guiding (prospective) students in their educational paths (Fonteyne, 2017; Fonteyne et al., 2017). GULS is a component of the wider SIMON test battery, administered to first-year students at the beginning of each academic year (i.e., October).

Construct Validity and Reliability

For the construct validity and reliability analyses, we used two samples. Only students with complete responses to the GULS, sex, and socio-economic status (SES) items (for the operationalizations, see *Measures*) in the SIMON tool were incorporated in both samples. Sample 1 comprised N_1 = 12,527 first-year higher education students from 2017-2018 to 2019-2020 (25% low SES; 60% female), and sample 2 included N_2 = 17,204 first-year higher education students from 2020-2021 to 2022-2023 (22% low SES; 59% female). Sample 1 was used to refine the GULS item set, which subsequently served as the basis for the construct validity and reliability analyses conducted on sample 2.

Predictive Validity

For the predictive validity analyses, we employed a subsample from each of the two original samples. Specifically, we included only the students enrolled in study programs for which we had at least 50 total responses for all six academic years to ensure sufficient statistical power. Here, total responses refer to complete responses on GULS and the control variables (see below), along with recorded GPA and study success scores after the second chance exam period (i.e., August/September). As such, subsample 1 consisted of n_1 = 8,244 first-year higher education students from 2017-2018 to 2019-2020, and subsample 2 comprised n_2 = 10,891 first-year higher education students from 2020-2021 to 2022-2023, both across sixteen study programs. Table 1 outlines these subsample sizes categorized by sex, SES, and the mathematics category of the students' enrolled study program. For the subsample sizes per study program, see *Appendix A*, Table A1.

Table 1

Subsample Sizes by Sex, SES, and Mathematics Category

	Subsample 1	Subsample 2	Total
Sex			
Male	3,048	4,208	7,256
Female	5,196	6,683	11,879
SES			
Low SES	2,085	2,363	4,448
High SES	6,159	8,528	14,687
Mathematics Category			
Basic	4,847	6,487	11,334
Advanced	3,397	4,404	7,801

Note. Subsamples 1 and 2 respectively encompass students from academic years 2017-2018 to 2019-2020 and 2020-2021 to 2022-2023. The advanced mathematics study programs are associated with more advanced mathematical curricula. Through the SIMON project (Fonteyne, 2017), students enrolling in basic mathematics study programs complete a less difficult math assessment compared to students enrolling in more advanced mathematics study programs. For more information on the predictive validity of this basic mathematics test, see Fonteyne and colleagues (2015).

Measures

Academic Achievement

On the one hand, academic achievement as the dependent variable in the present study was operationalized by using students' *Grade Point Average (GPA)*. At Ghent University, students have two first-chance exam periods, one per semester. Each course is graded from 0 to 20, with 10 as the passing mark. After summer break, a second chance exam period is provided for students who did not pass one or more courses on their first attempt. Every course is also assigned a number of ECTS credits (European Credit Transfer and Accumulation System credits) (European Commission, 2015), which reflect its weight and are included in the GPA calculation. In the present study, GPA was scored from 0 to 100.

On the other hand, *study success* was also used as an academic achievement measure. A first-year student in the standard curriculum can accumulate a maximum of 60 ECTS credits. Study success was defined as the ratio of students' obtained ECTS credits over their subscribed ones after the second chance exam period and ranged between 0 and 100.

Language Proficiency

Language Proficiency was operationalized by using GULS, which constitutes the central focus of the present study. The aim of developing GULS was to integrate the test into the existing SIMON tool for (prospective) higher education students, focusing on study (re)orientation and remediation (Fonteyne, 2017; Fonteyne et al., 2017). Recognizing the need for a comprehensive language proficiency test beyond vocabulary, developers piloted the initial 45-item GULS with last-year secondary students in 2016-2017. A team of experts, including linguists and experimental and cognitive psychologists, developed and (re)evaluated the items. To optimize student testing time and based on preliminary analyses that showed promising outcomes regarding the total scores' descriptives, reliability, and item difficulty and discrimination parameters, the developers subsequently shortened GULS to 25 items. For further details, refer to *Appendix C*, Preliminary Analyses Pilot GULS, and *Appendix D*, Items GULS. In the present study, we therefore assessed the construct validity and reliability of GULS using data from the academic years 2017-2018 to 2019-2020, resulting in a refined set of GULS items. This selected set of GULS items served as the basis for the subsequent construct/predictive validity and reliability analyses on the data from the academic years 2020-2021 to 2022-2023 (see also *Analyses*).

Control Variables

Sex distinguished between males (= 0) and females (= 1) as stated on the student's passport.

Socio-Economic Status (*SES*) considered a student's scholarship status and their parents' educational attainment. Students are categorized into the low-SES group (= 0) versus the high-SES group (= 1) if neither parent completed secondary education and/or if they receive a scholarship. Such an operationalization of SES is recognized as a relevant predictor of academic achievement in previous research (Fonteyne, 2017; Hauspie et al., 2024).

Mathematics Hours Secondary Education refers to the students' self-reported total number of hours mathematics instruction they received during their last year of secondary education. Similar to the personal background characteristic SES, this educational background characteristic is an established predictor of academic achievement (Fonteyne, 2017; Schelfhout, 2019).

Vocabulary Knowledge was measured through the LexTALE test (Lemhöfer & Broersma, 2012). Students were asked to indicate, for each of the 60 items (ω's > .70), whether the item represented an actual word or not.

Mathematics Category refers to the category in which the studied study programs were classified: basic (= 0) and advanced (= 1), with the latter featuring more advanced mathematical curricula. This control variable was specifically used in the analyses regarding GULS's predictive validity for academic achievement on the 2017-2018 to 2019-2020 and 2020-2021 to 2022-2023 data across study programs.

Mathematical Proficiency was measured using either the basic or advanced mathematics test within the SIMON tool (Fonteyne, 2017), depending on the student's enrolled study program (see also *Mathematics Category*). The basic mathematics test comprised 20 items (open and multiple-choice; ω's > .70), while the advanced one consisted of 25 multiple-choice items (ω's > .70) and was more challenging. For examples and more information on the predictive validity of the basic mathematics test, see Fonteyne (2017) and Fonteyne and colleagues (2015). This control variable was specifically used in the analyses regarding GULS's predictive validity for academic achievement on the 2017-2018 to 2022-2023 data per study program.

Analyses

Construct Validity and Reliability

First, we evaluated the construct validity and reliability of GULS on the data from 2017-2018 to 2019-2020 (i.e., sample 1), with 25 GULS items. A Univariate Variable Analysis (UVA) was conducted to assess local independence. This analysis identifies items with a weighted Topological Overlap (wTO) exceeding .25, suggesting potential interdependence and recommending removal of one item to mitigate redundancy (Christensen et al., 2023). Building upon literature on language proficiency, we proceeded with a Confirmatory Factor Analysis (CFA) employing a single factor. We used the Weighted Least Squares Mean and Variance adjusted (WLSMV) estimator due to the categorical nature of our GULS data (Rosseel, 2012). The robust fit indices were examined to evaluate the model fit (i.e., chi-square test, Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR)). A non-significant chi-square test, CFI above .90, and RMSEA and SRMR values below .06 indicate good fit. RMSEA and SRMR values between .06 and .08 signify acceptable fit (Rosseel, 2012). Additionally, we scrutinized the factor loadings (λ) of the items, excluding those with $\lambda \leq .30$ (Tavakol & Wetzel, 2020) and reevaluating the model. Subsequently, McDonald's ω was employed to assess the reliability of the remaining item set after the UVA and CFA analyses, following the commonly used threshold that $\omega \geq .70$ indicates acceptable reliability (Dunn et al., 2014). Moreover, $.60 \leq \omega < .70$ is considered sufficient for decision making at group level, and $.70 \leq \omega < .80$ at individual level (Evers et al., 2009; Nunnally & Bernstein, 1994). Before conducting the Item Response Theory (IRT) analyses, we additionally examined the unidimensionality of the retained GULS item set following the UVA and CFA analyses. A unidimensional bifactor model was tested, including one general factor capturing shared variance and three group factors capturing residual item clustering. The inclusion of three group factors instead of two provide a more conservative and robust check for potential multidimensionality. The general factor is considered dominant if omega hierarchical (ω_h) $\geq .50$ and explained common variance (*ECV*) $\geq .50$. Low omega values for group factors ($\omega \leq .20$) indicate minimal residual multidimensionality (Reise et al., 2013; Reise et al., 2023).

Further, we applied IRT analyses to the dichotomous data concerning the retained GULS items. More specifically, we compared the model fits of both the one- and two-parameter IRT models, along with the IRT model incorporating a guessing parameter, using Analysis of Variance tests. The Test Information Curves (TICs) of these IRT models were plotted with a particular focus on clarifying the amount of information provided by the test within the lower range [-6, 0]. At the item level, we calculated the item-total score correlations and examined the item in-and outfit statistics[3], aiming for values between the traditional mean square item fit bounds of 0.75 and 1.33 (Katz et al., 2021). Subsequently, we classified the item difficulty parameters (b) as follows: $b < -2.00$ (very easy), $-2.00 \leq b \leq 2.00$ (moderately difficult), and $b > 2.00$ (very difficult) (Hambleton et al., 1991). The item discrimination parameters (a) were categorized in the following way: $0.01 \leq a \leq 0.34$ (very low), $0.35 \leq a \leq 0.64$ (low), $0.65 \leq a \leq 1.34$ (moderate), $1.35 \leq a \leq 1.69$ (high), $a \geq 1.70$ (very high) (Baker, 2001). Corresponding visualizations were operationalized through Item Characteristic Curves (ICCs).

Second, this data analysis procedure (i.e., UVA, CFA, ω, IRT with TICs, and ICCs) was reiterated on the 2020-2021 to 2022-2023 data (i.e., sample 2), using the recommended set of GULS items derived from the analyses of the preceding three years.

[3] Item infit and outfit assess the fit between observed and expected item responses in IRT models, with item infit focusing on responses near the item's difficulty and item outfit considering overall fit across ability levels. Item outfit is also sensitive to outliers.

Predictive Validity

First, we examined the predictive validity of GULS for academic achievement *across study programs* on the 2017-2018 to 2019-2020 (i.e., subsample 1) and 2020-2021 to 2022-2023 data (i.e., subsample 2) separately (see also *Participants*). We controlled for sex, SES, mathematics hours secondary education, vocabulary knowledge, and mathematics category. After checking the sample-specific reliability of GULS using McDonald's *w*, GULS was included as a total score (i.e., the mean of the retained set of GULS items after the subsample-specific construct validity and reliability analyses). For both subsamples, the descriptives and correlation matrix were examined, followed by a linear regression with academic achievement (i.e., GPA or study success) as the dependent variable and GULS, along with the control variables, as the predictors. The Pearson correlation coefficient was used with the following rule of thumb: $|r| = .00$ (no correlation), between $.00 < |r| < .20$ (very weak), $.20 \leq |r| < .40$ (weak), $.40 \leq |r| < .60$ (moderately strong), $.60 \leq |r| < .80$ (strong), $.80 \leq |r| < 1.00$ (very strong), and $|r| = 1.00$ (perfect) (Krehbiel, 2004). Further, we focused on GULS's parameter estimates, and its individual and unique explained variance in academic achievement using Adjusted R-squared. Individual explained variance indicates a predictor's total impact on the dependent variable regardless of other predictors, while unique explained variance shows its distinct contribution (i.e., incremental predictive validity). Also, we investigated whether the explained variance in academic achievement differed between the full model and the model without GULS. Multicollinearity was assessed using Variance Inflation Factor (VIF) values for each independent variable. Roughly, VIF values below 10 are acceptable, though values over 5 may indicate substantial multicollinearity, suggesting a threshold of VIF < 5 for reliable regression results (Marcoulides & Raykov, 2019).

Second, we investigated the predictive validity of GULS for academic achievement *per study program* (segmented by math category). We controlled for sex, SES, mathematics hours secondary education, vocabulary knowledge, and mathematical proficiency. To ensure sufficient statistical power per study program, we used the total dataset from 2017-2018 to 2022-2023 (see also *Participants*). Per study program, reliability was verified through McDonald's ω. Descriptives were computed, and linear regressions were conducted with GPA or study success as the dependent variable, and GULS and the control variables as the independent variables. Again, focus was directed towards GULS's parameter estimates, its individual and unique explained variances in academic achievement using Adjusted R-squared, and the difference in explained variance in academic achievement between the full model and model without GULS. Multicollinearity was also tested per study program (Marcoulides & Raykov, 2019).

Results

Construct Validity and Reliability

Sample 1

In the dataset spanning from 2017-2018 to 2019-2020, we initially considered 25 GULS items. Upon evaluating their wTO values in relation to other items, the UVA shows the exclusion of GULS2 (wTO = .26 paired with GULS1), GULS21 (wTO = .47 paired with GULS25), and GULS23 (wTO = .38 paired with GULS22). The initial CFA on the remaining 22 GULS items reveals a good model fit ($\chi^2_{scaled}(209) = 1{,}728.53$, $p < .001$, $CFI_{robust} = .901$, $RMSEA_{robust} = .049$, $SRMR = .039$). After removing GULS19 ($\lambda = .24$) and GULS20 ($\lambda = .30$), the second CFA again demonstrates a good model fit ($\chi^2_{scaled}(170) = 947.69$, $p < .001$, $CFI_{robust} = .926$, $RMSEA_{robust} = .045$, $SRMR = .034$) but highlights the exclusion of GULS11 ($\lambda = .29$) and GULS18 ($\lambda = .29$). The final CFA confirms the good model fit, with all 18 GULS items showing factor loadings $\lambda > .30$. See Table 2 for the final model fit and reliability measures, and Table 3 for the item factor loadings.

The unidimensional bifactor model on the 18-item GULS set shows that the general factor dominates, with ω_h = .62 and *ECV* = .58. Omega values for the group factors range from ω = .10 to ω = .16), indicating minimal multidimensionality.

In the IRT analysis, the two-parameter model (AIC = 189,380.00) with 18 GULS items (*M* = 14.28, *SD* = 2.83) outperforms the one-parameter model (AIC = 190,576.30, $p < .001$). The three-parameter model with guessing parameter (AIC = 189,318.10, $p < .001$) demonstrates a better model fit compared with the two-parameter model. However, the two-parameter model offers more test information within the lower range of language proficiency ability [-6, 0] (85% versus 82% attained by the three-parameter model). See Figure 1 for the TIC of the two-parameter model, and *Appendix A,* Figure A1 for the TIC of the three-parameter model. At the item level, all item in-and outfit statistics meet the accepted criteria, falling between 0.75 and 1.33. Item difficulties range from -2.84 to 0.33, and item discriminations vary between 0.62 and 2.11. ICCs are available in *Appendix A*, Figure A2. Plots regarding the item in- and outfits can be found in Figure A3. The item parameters and item-total score correlations are displayed in Table 3.

Sample 2

In the dataset spanning from 2020-2021 to 2022-2023, we considered 18 GULS items. This 18-item set remains consistent following UVA and CFA. No wTO > .25 values are found between two items, the model fit is good with none of the item factor loadings are $\lambda \leq .30$. Refer to Table 2 for the model fit and reliability measures, and Table 3 for the item factor loadings.

The unidimensional bifactor model on the 18-item GULS set indicates a dominance of the general factor (ω_h = .68; *ECV* = .75). Multidimensionality among the group factors is minimal, with omega values ranging from ω = .07 to ω = .19.

In the IRT analysis, the two-parameter model (AIC = 274,752.50) using 18 GULS items (M = 13.86, SD = 3.07) demonstrates superior performance compared to the one-parameter model (AIC = 276,750.50, $p < .001$). The three-parameter model (AIC = 274,628.30) incorporating a guessing parameter shows a better model fit compared with the two-parameter model ($p < .001$). However, the two-parameter model provides more test information within the lower range of language proficiency ability [-6, 0] (84% versus 81% obtained by the three-parameter model). Refer to Figure 1 for the TIC of the two-parameter model, and to *Appendix A,* Figure A4 for the TIC of the three-parameter model. At the item level, all in-fit and outfit statistics align with the target range of 0.75 to 1.33. Item difficulties span from -2.26 to 0.43, while item discriminations range between 0.62 and 2.17. The ICCs can be found in *Appendix A*, Figure A5. Plots regarding the item in- and outfits can be found in Figure A6. The item parameters and their total score correlations are presented in Table 3.

Table 2

Model Fit and Reliability per Sample

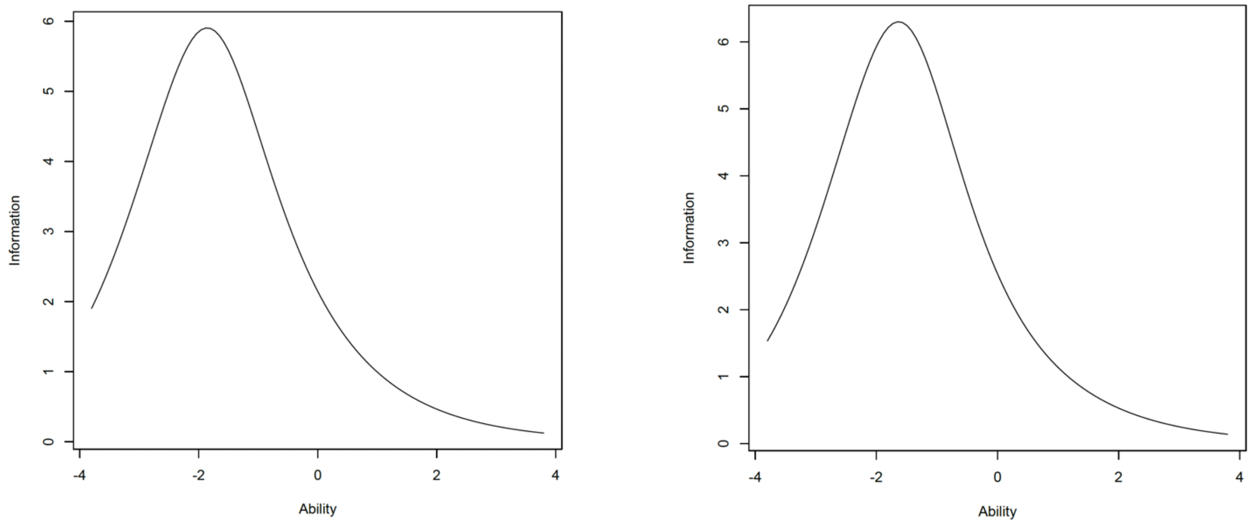
	Sample 1	Sample 2
χ^2_{scaled} (135) (p)	684.20 (< .001)	930.44 (< .001)
CFI_{robust}	.936	.953
$RMSEA_{robust}$	.047	.041
SRMR	.032	.029
McDonald's ω	.74	.76

Note. χ^2_{scaled} (df) = chi-square test with degrees of freedom placed in parentheses, CFI_{robust} = Robust Comparative Fit Index, $RMSEA_{robust}$ = Robust Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual. The first four rows present the results for the goodness-of-fit measures. The last row pertains to the reliability results. Samples 1 and 2 respectively encompass students from academic years 2017-2018 to 2019-2020 (N_1 = 12,527) and 2020-2021 to 2022-2023 (N_2 = 17,204). The reported values correspond to the 18-item final model.

Figure 1

Test Information Curves Two-Parameter Models per Sample

Sample 1 Sample 2



Note. N_1 = 12,527, N_2 = 17,204. The two-parameter model of sample 1 offers 85% test information within the lower range of language proficiency ability [-6, 0], while the two-parameter model of sample 2 offers 84%.

Predictive Validity

For the predictive validity analyses across study programs and study program-specific, no VIF value exceeds 2, indicating the absence of multicollinearity issues. We refer to *Appendix A*, Figure A7, for the scatter plots. The predictive validity results of language proficiency for study success (after the second chance exam period), across study programs and study-program specific, can be found in *Appendix B*, Predictive Validity Study Success After Second Chance Exam Period.

Across Study Programs

Subsample 1. For the 2017-2018 to 2019-2020 data across sixteen study programs, the correlation matrix can be found in *Appendix A*, Table A3. GULS (M = 14.1, SD = 2.9) has a significant effect on academic achievement (i.e., GPA), when controlled for sex, SES, mathematics hours secondary education, vocabulary knowledge, and mathematics category ($t(8{,}237) = 18.20$, $p < .001$, $B = 1.20$, $\beta = 0.20$). Individually, the explained variance of GULS for GPA is $R^2 = .06$. GULS's unique contribution to the explained variance in GPA is $R^2 = .04$. The difference in explained variance between the full model ($R^2 = .12$) and the model without GULS is significant ($F_{Change}(1, 8{,}237) = 331.32$, $p < .001$). The reliability $w = .73$.

Subsample 2. For the 2020-2021 to 2022-2023 data across sixteen study programs, the correlation matrix is available in *Appendix A*, Table A3. GULS ($M = 13.6$, $SD = 3.1$) has a significant effect on GPA, when controlled for sex, SES, mathematics hours secondary education, vocabulary knowledge, and mathematics category ($t(10{,}884) = 23.87$, $p < .001$, $B = 1.33$, $\beta = 0.22$). Individually, the explained variance of GULS for GPA is $R^2 = .08$. GULS's unique contribution to the explained variance in GPA is $R^2 = .04$. The difference in explained variance between the full model ($R^2 =.17$) and the model without GULS is significant ($F_{Change}(1, 10{,}884) = 569.85$, $p < .001$). The reliability $w = .75$.

Study Program-Specific

The data from 2017-2018 to 2022-2023 show that GULS significantly predicts GPA for all the basic mathematics study programs (*M* range [13.2 - 15.2], *SD* range [2.2 - 3.1]), when controlled for sex, SES, mathematics hours secondary education, vocabulary knowledge and mathematical proficiency (i.e., for each study program, $p < .001$). Individually, the explained variance of GULS for GPA varies from $R^2 = .07$ (Physical Therapy and Motor Rehabilitation, and Political Sciences) to $R^2 =.21$ (Applied Language Studies). GULS's unique contribution to the explained variance in GPA ranges from $R^2 = .02$ (Physical Therapy and Motor Rehabilitation) to $R^2 = .11$ (Applied Language Studies). The differences in explained variance between the full models and the models without GULS are significant (*p*'s < .001).

Across the advanced mathematics study programs (*M* range [12.7 - 14.9], *SD* range [2.7 - 3.3]), the study programs Biochemistry and Biotechnology, and Engineering Technology diverge from the observed pattern where GULS significantly predicts GPA, when controlled for sex, SES, mathematics hours secondary education, vocabulary knowledge and mathematical proficiency (*p* range [< .001 - .260]). Individually, the explained variance of GULS for GPA varies R^2 = .02 (Engineering Technology) to R^2 = .08 (Business Administration and (Applied/Business) Economics). GULS's unique contribution to the explained variance in GPA ranges from R^2 < .01 (Biochemistry and Biotechnology, and Engineering Technology) to R^2 = .05 (Business Administration). The differences in explained variance between the full models and the models without GULS are significant, except for the study programs Biochemistry and Biotechnology, and Engineering Technology (*p* range [< .001 - .260]). For details on GULS's descriptives, predictive validity measures for GPA, reliability, and the total explained variance of the full model per study program, we refer to Table 4.

Table 3

Item Factor Loadings, Difficulty and Discrimination Parameters and Item-Total Correlations per Sample

Item	Factor Loading		Item Difficulty		Item Discrimination		Item-Total Correlation	
	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2
GULS1	.59	.57	-2.28	-2.06	1.33	1.22	.32	.34
GULS3	.52	.51	-0.11	0.20	1.01	0.99	.33	.36
GULS4	.36	.37	0.33	0.43	0.62	0.64	.22	.24
GULS5	.40	.42	-1.60	-1.37	0.75	0.79	.26	.29
GULS6	.54	.50	-1.53	-1.52	1.11	1.01	.33	.33
GULS7	.65	.68	-2.20	-2.05	1.53	1.59	.33	.37
GULS8	.35	.33	0.01	0.28	0.65	0.62	.23	.23
GULS9	.59	.60	-1.95	-1.87	1.27	1.30	.34	.36
GULS10	.41	.43	-1.96	-1.96	0.76	0.79	.26	.27
GULS12	.55	.55	-1.95	-1.79	1.14	1.14	.31	.33
GULS13	.57	.56	-1.58	-1.52	1.19	1.16	.34	.35
GULS14	.52	.55	-1.57	-1.42	1.08	1.15	.34	.36
GULS15	.49	.58	-2.84	-2.26	1.04	1.23	.24	.31
GULS16	.59	.66	-2.49	-1.91	1.35	1.53	.29	.37
GULS17	.50	.50	-1.84	-1.80	1.00	0.97	.31	.31
GULS22	.67	.68	-1.45	-1.14	1.52	1.59	.43	.46
GULS24	.77	.78	-1.79	-1.52	2.11	2.17	.43	.48
GULS25	.51	.56	-1.92	-1.59	1.03	1.17	.33	.38

Note. Samples 1 and 2 respectively encompass students from academic years 2017-2018 to 2019-2020 (N_1 = 12,527) and 2020-2021 to 2022-2023 (N_2 = 17,204). Exclusions after UVA comprised GULS2, GULS21, and GULS22. Following CFA's, exclusions involved GULS11, GULS18, GULS19, and GULS20. Further details are available in *Method*.

Table 4

GULS's Descriptives, Predictive Validity Measures for GPA, and Reliability, and Total Explained Variance Full Model per Study Program

	M (SD)	t	B (b)	R^2			ω
				Individual	Unique	Total	
Basic mathematics							
Applied Language Studies	14.7 (2.6)	9.11***	2.79 (0.31)	.21	.11***	.28	.70
Communication Sciences	14.2 (2.6)	4.63***	1.38 (0.30)	.09	.04***	.13	.68
Criminological Sciences	13.5 (2.8)	7.50***	1.37 (0.18)	.09	.04***	.16	.68
Educational Sciences	13.9 (2.6)	7.69***	1.38 (0.18)	.11	.08***	.19	.63
History	15.2 (2.2)	4.12***	1.94 (0.47)	.12	.03***	.18	.64
Law	14.3 (2.7)	13.28***	1.87 (0.14)	.15	.07***	.22	.72
Pharmaceutical Sciences	13.7 (3.1)	6.58***	1.19 (0.18)	.11	.03***	.19	.77
Physical Therapy and Motor Rehabilitation	13.2 (3.1)	6.46***	0.80 (0.12)	.07	.02***	.20	.74
Political Sciences	14.5 (2.5)	4.55***	1.57 (0.35)	.07	.04***	.10	.68
Psychology	13.8 (2.8)	12.73***	1.55 (0.12)	.11	.05***	.22	.69
Advanced mathematics							
Biochemistry and Biotechnology	14.1 (3.1)	1.18	0.34 (0.29)	.03	< .01	.14	.78
Biomedical Sciences	13.7 (3.1)	4.10***	0.82 (0.20)	.06	.01***	.14	.76
Bioscience Engineering	14.9 (2.7)	3.29**	0.60 (0.18)	.07	.01**	.24	.78
Business Administration	12.7 (3.3)	10.43***	1.25 (0.12)	.08	.05***	.15	.75
(Applied/Business) Economics	13.9 (3.1)	7.70***	0.98 (0.13)	.08	.03***	.17	.78
Engineering Technology	13.4 (3.2)	1.13	0.17 (0.15)	.02	< .01	.16	.77

Note. $*p < .05$, $**p < .01$, $***p < .001$. ω = McDonald's ω as reliability measure. R^2 = Adjusted R-squared. Total R^2 represents the explained variance by the full model, whereas unique R^2 quantifies the specific contribution of GULS to GPA, distinct from other predictors (i.e., sex, SES, mathematics hours secondary education, vocabulary knowledge, and mathematical proficiency). Meanwhile, individual R^2 reflects the overall contribution of GULS to GPA, regardless of additional predictors. The advanced mathematics study programs are associated with more advanced mathematical curricula. Through the SIMON project (Fonteyne, 2017), students enrolling in basic mathematics study programs complete a less difficult math assessment compared to students enrolling in more advanced mathematics study programs.

Discussion

Academic language proficiency is particularly important for students entering higher education (Knoch & Elder, 2013; Read, 2016). Indeed, language proficiency, assessed for admission eligibility (e.g., Clinton-Lisell et al., 2022; Ihlenfeldt & Rios, 2023) or through post-entry language assessment (PELA) tests (e.g., Heeren et al., 2021; van Dijk, 2015), shows predictive validity for academic achievement (Elder, 2017; Knoch & Elder, 2013).

In the absence of crucial information about the cognitive and academic skills of incoming students, PELAs are notably relevant for higher education institutions with low or no admission requirements. These assessments help identify at-risk students and provide them with the necessary language support to improve their academic trajectory (Elder, 2017; Knoch & Elder, 2013; Read, 2016). However, few PELAs are available for languages other than English like Dutch. An existing Dutch PELA, developed by Heeren and colleagues (2021) certainly has merits, but the test and its detailed construct validation evidence are not publicly accessible. The researchers also used a less conventional, however still valuable, metric of academic achievement compared to GPA to examine the predictive validity of language proficiency for academic achievement. Moreover, the differential contribution of language proficiency to academic achievement for various academic disciplines is acknowledged (Elder, 2017; Read, 2016), yet often not implemented.

The present study aimed to enhance PELA in higher education by introducing and validating the Ghent University Language Screening (GULS) as a fully open access (i.e., easy-to-administer, free and publicly available) PELA. Concretely, GULS assesses Dutch language proficiency, in terms of reading comprehension, of first-year higher education students. First, GULS's construct validity and reliability were evaluated and confirmed through data from the academic years 2017-2018 to 2019-2020 (N_1 = 12,527) and 2020-2021 to 2022-2023 (N_2 = 17,204). Second, GULS’s predictive validity for academic achievement (i.e., GPA and study success) was assessed on data over these two three-year periods across sixteen study programs (n_1 = 8,244; n_2 = 10,891), succeeded by predictive validity analyses for each study program across the combined six-year span (*n*'s > 400).

Construct Validity and Reliability

The results concerning construct validity indicate that GULS with 18 items is the optimal version to avoid item redundancy (Christensen et al., 2023) and ensure sufficiently high item factor loadings (Tavakol & Wetzel, 2020). Data from the periods 2017-2018 to 2019-2020 and 2020-2021 to 2022-2023 subsequently fit well with the one-factor model comprising 18 items, where the factor represents language proficiency in terms of reading comprehension (Kintsch, 2013; OECD, 2023). Additionally, GULS appears to provide the most test information at the lower end of the language proficiency spectrum. In other words, GULS is particularly sensitive and insightful for students with lower language proficiency scores, enabling the identification and assistance of those students in need of additional language development support. This finding corresponds with the objectives of PELAs (Elder, 2017; Knoch & Elder, 2013; Read, 2016). Especially in low-stakes higher education environments with a more diverse enrolling student population, the ability to detect and support such at-risk students at an early stage is key to be able to achieve the so-called equity for equal achievement (Espinoza, 2007).

GULS's reliabilities across study programs exceed the rule of thumb. Hence, GULS serves multiple purposes effectively (Evers et al., 2009; Nunnally & Bernstein, 1994). First, the test can be used in population-level research across and within higher education study programs, including both basic as well as advanced mathematics study programs. Second, non-binding advice can be offered to (prospective) first-year higher education students regarding their language proficiency in terms of reading comprehension. Third, tailoring this individual advice to specific advanced mathematics study programs is possible based on

GULS's reliabilities. However, caution is warranted for basic mathematics study programs, as the empirical evidence regarding the reliabilities is inconclusive as such in six out of the 10 included study programs (although, they do satisfy the threshold for population-level research) (Evers et al., 2009; Nunnally & Bernstein, 1994).

Moreover, the construct validity and reliability findings across study programs are confirmed by a second, independent prospective dataset from 2021-2022 to 2022-2023. This fact reinforces the robustness of GULS as a valid and reliable tool for assessing Dutch language proficiency, in terms of reading comprehension, of first-year students in higher education.

Predictive Validity

Consistent across the datasets from 2017-2018 to 2019-2020 and from 2020-2021 to 2022-2023, GULS is a significant predictor of academic achievement (i.e., GPA and study success) across sixteen study programs. Language proficiency measured through GULS contributes modestly to the prediction of academic achievement, aligning with the correlations found in previous research using language assessment for admission in English-medium higher education institutions (for meta-analyses, see e.g., Ihlenfeldt & Rios, 2023; Clinton-Lisell et al., 2022) or (non-)Dutch PELAs (e.g., De Wachter et al., 2013; Heeren et al., 2021; van Dijk, 2015). Researchers also emphasize that language proficiency typically explains no more than 10% of the variance in academic achievement (Elder, 2017; Knoch & Elder, 2013). We further acknowledge that language proficiency is a relevant yet not exclusive determinant of academic achievement. The interplay between (non-)cognitive and personal/educational background characteristics is indeed essential for accurately predicting academic achievement (Elder, 2017; Fonteyne et al., 2017; Richardson et al., 2012; Read, 2016). Importantly, even when controlling for personal and educational background characteristics, vocabulary knowledge, and mathematics category in the present study, we observe incremental predictive validity of GULS for academic achievement. This finding implies that language proficiency assessed through GULS continues to capture unique aspects of and provide insights into academic achievement that are not entirely covered by other well-established predictors for academic achievement like SES. Literature indicates that personal background characteristics rather weakly contribute to academic achievement beyond language proficiency and do not impact the incremental predictive validity of language proficiency, unlike educational background characteristics (Heeren et al., 2021). In the present study, accounting for mathematics hours in secondary education as an educational background characteristic also increases the total variance explained in GPA across study programs. At the same time, however, this educational background characteristic does not alter the unique contribution of language proficiency as measured by GULS to academic achievement.[4] Researchers' observation that educational background characteristics are primarily reflected in language proficiency (Heeren et al., 2021; Stricker, 2004) does not emerge prominently in our context across study programs.

Beyond the primary linear analyses, we additionally examined whether the association between GULS and academic achievement varies across the score range using Generalized Additive Models (GAMs). The analyses show that the association is most pronounced at the lower end of the GULS scale (scores ≤ 10), with lower GULS scores linked to progressively lower predicted academic achievement relative to the sample mean. At GULS = 10, the negative smooth effect is still larger in absolute value than the positive effect observed at the highest GULS scores, indicating that low GULS scores have a greater predicted impact on academic achievement than high scores. Thereafter, the effect gradually attenuates and

[4] The difference in unique explained variance of GULS with or without mathematics hours secondary education in the regression model (including the other control variables) amounts 1% for both GPA as well as study success, on average across the two three-year periods.

becomes positive from approximately GULS = 15 onward. Confidence intervals are wider for the lowest GULS scores, reflecting greater uncertainty due to fewer observations at these extremes, and findings for very low GULS scores should therefore be interpreted with caution. A similar pattern is observed for study success, although the effects span a wider range. GAM plots and results can be found in *Appendix A* (Figure A8; Tables A4–A5). Taken together, these findings suggest that GULS may be particularly useful for identifying students at elevated academic risk, especially those with low GULS scores. This pattern can also help to explain why linear models account for only a modest proportion of variance and aligns with test information results indicating greater measurement precision at lower levels of GULS.

Further, study-program specific analyses spanning the six academic years likewise indicate that GULS modestly predicts academic achievement across all basic mathematics study programs. Similarly to the overall analysis, after accounting for personal and educational background characteristics, vocabulary knowledge, and mathematical proficiency, GULS also shows incremental predictive validity for academic achievement. The models including GULS are regarded superior to the models without GULS for predicting academic achievement. Notably, the role of language proficiency in academic achievement prediction may differ by discipline (Elder, 2017; Knoch & Elder, 2013; Read, 2016), regardless of the comparable level of mathematics proficiency that study programs expect from enrolling students. Indeed, the present study confirms a differential contribution of GULS to academic achievement. For example, the Applied Language Studies study program stands out in terms of GULS's predictive validity for academic achievement, with an individual and unique contribution of 21% and 11%, respectively. Given the study program's focus on language, culture and its practical applications to real-world contexts (Berns & Matsuda, 2006), this outcome is not surprising and illustrates GULS's construct validity well. By contrast, in the discipline of Physical Therapy and Motor Rehabilitation within the basic mathematics study programs, GULS only accounts for 2% of the variance in GPA and 1% in study success when controlled for the other predictors. Yet, the inclusion of GULS still renders the model more effective for predicting academic achievement than the model without GULS. Furthermore, the educational background variable of mathematics hours in secondary education contributes to 7% of the unique variance in both GPA as well as study success within this specific study program. Nevertheless, the impact of this educational background variable on the unique explained variance of GULS is minimal, as this educational background variable only reduces the unique explained variance of GULS by 1% in both GPA and as well as study success when included in the model.

The same pattern regarding the modest (incremental) predictive validity of GULS for academic achievement applies to most advanced mathematics study programs as well. The control variable, mathematics category, indeed shows no effect on academic achievement when controlling for other predictors in the models across study programs. Overall, even for students who choose a study program that is characterized by more advanced mathematical curricula, language proficiency as measured through GULS remains important for their academic achievement. In fact, language and mathematical proficiency are established predictors of academic achievement (e.g., Fonteyne et al., 2017). Moreover, meta-analytic evidence shows a mutual moderate relationship between these proficiencies, primarily attributed to domain-general processes such as executive functions (Ünal et al., 2023). In addition, GULS proves to be sensitive to the role of language proficiency in predicting academic achievement, as its predictive validity for academic achievement also varies by discipline within the advanced mathematics study program group (Elder, 2017; Knoch & Elder, 2013; Read, 2016). However, for the advanced mathematics study programs Biochemistry and Biotechnology, and Engineering Technology, we find no effect of GULS on academic achievement. This result may be due to these study programs' reliance on more deeply entrenched technical and specific academic language, while GULS is developed to evaluate more general academic language proficiency. Looking further, we also observe that the mathematical proficiency in the study program Engineering Technology and the mathematics hours in secondary education in Biochemistry and Biotechnology explain the most in their prediction models for academic achievement, respectively accounting for 5% and 3% of the variance in GPA and for 5% and 2% of the variance in study success.

Implications

To the best of our knowledge, the present study addresses some gaps in current literature. First, the existence of GULS fills a void concerning the availability of Dutch PELAs, while also being fully open access (i.e., easy-to-administer, free, and publicly available). Second, we reported in detail on the construct validity and reliability, thereby enhancing transparency, replicability, and reproducibility (Min & Aryadoust, 2021), and facilitating public access to the psychometric properties of a PELA (Knoch & Elder, 2013). Third, we examined GULS's (incremental) predictive validity for academic achievement both across as well as within various study programs (Elder, 2017; Hauspie et al., 2024; Read, 2016). As a result, the present study provides insights into the overall and differential contribution(s) of language proficiency to academic achievement, including linguistic, social science, economic, and STEM-oriented disciplines, among others.

The present study validated GULS as a Dutch PELA that offers practical benefits due to its open accessibility, enabling its ease of deployment in research and practice contexts. With demonstrated construct validity and reliability, GULS proves useful for conducting population-level research across and within diverse study programs (Evers et al., 2009; Nunnally & Bernstein, 1994). For instance, GULS can be utilized to control for Dutch language proficiency when examining academic achievement in future educational research.

Moreover, GULS facilitates the provision of individualized, non-binding advice to (prospective) first-year higher education students regarding their Dutch language proficiency levels. This personalized guidance can help identify students at heightened academic risk. Based on our analyses, students with lower GULS scores tend to face greater academic difficulties, with the associated risk gradually decreasing as scores increase. Within this risk gradient, a GULS score of 10 emerges as a pragmatic, model-informed reference point. Specifically, this is the highest score at which the negative association with academic achievement still exceeds the maximal positive association observed at the upper end of the scale in absolute value. In the present study, this score also corresponds to the lower tail of the distribution (approximately 13.2% of students), supporting its use as a conservative threshold for identifying at-risk students. Importantly, this reference point should not be interpreted as a strict cut-off. Risk may vary across study programs, and uncertainty is greater for very low GULS scores. Therefore, GULS = 10 should be seen as a guideline within a continuous risk gradient, rather than an absolute boundary. While this score can meaningfully inform support strategies, researchers and practitioners are encouraged to evaluate its functioning within their own samples to ensure appropriate interpretation and application in specific institutional contexts.

Building on this identification of at-risk students, short, targeted language support interventions can be offered to improve reading comprehension (Pyle et al., 2017; Sohn et al., 2023). The implications of such interventions for academic performance, however, should be interpreted with caution. Even though the results of our study suggest an influence of reading comprehension on academic achievement, supported by temporal precedence and prediction beyond key covariates, it remains uncertain whether improving reading comprehension would also translate into higher GPA and/or study success. Determining the impact of such interventions on academic achievement requires further research.

Notably, GULS also allows for the specification of advice for advanced mathematics study programs, although caution is advised for basic mathematical programs (Evers et al., 2009; Nunnally & Bernstein, 1994). In any case, it is always encouraged to verify reliability within one's own sample in new research (Graham, 2015; Harris, 2003).

Strengths, Limitations and Future Research

GULS assesses language proficiency in terms of reading comprehension, beyond simple vocabulary knowledge, to identify at-risk students, offer language support and enrich their academic journey. We opted to focus on reading comprehension, given its paramount importance during the early stages of higher education (De Wachter et al., 2013; Jansen et al., 2022; Van Houtven et al., 2010). Consequently, GULS does not involve other language proficiency skills such as writing, which can be interesting to address in future research and application.

For the construct/predictive validity and reliability analyses across study programs, large sample sizes were used, with initial results from a three-year period subsequently confirmed by data from an additional three years across study programs. To ensure robust statistical power for the program-specific predictive validity analyses, we utilized the student data over the six years. On this matter, to examine the predictive validity of GULS for academic achievement, we considered both GPA and study success as academic achievement measures. As such, we address operationalizations of academic achievement that are applicable to both European (i.e., study success) and American educational contexts (i.e., GPA). Future (longitudinal) research could also investigate GULS's predictive validity for timely bachelor's degree completion, allowing for a more comprehensive assessment of how language proficiency impacts students' academic progress and degree attainment over time.

In addition, we accounted for a variety of control variables in the predictive validity analyses, including personal and educational background characteristics and cognitive factors.

Non-cognitive factors, however, were not considered in the present study, highlighting a direction for future research as much of the variance in academic achievement remains unexplained. Moreover, attention could also be given to (theory-based) interactions between language proficiency and various other variables. Educational background characteristics like data on standardized exams were non-existent, while objective data on high school GPA were not available. The number of mathematics hours in secondary education was therefore regarded as a more objective indicator of educational background, although students' answers also depended on self-reporting.

Conclusion

In the present study we validated GULS, a Dutch post-entry language assessment with fully open access. More specifically, GULS assesses reading comprehension of first-year students in higher education. GULS is proven to be valid and reliable, particularly for identifying students who require language support at the start of higher education, and predicts academic achievement. The accessibility of GULS enhances its utility as a Dutch language proficiency test for future population-level research and for providing advice to (prospective) higher education students.

Declarations

Funding

The present study was funded by Ghent University through the SIMON project (WBS Element: E/00950/01).

Conflict of interest/Competing interests

The authors have no relevant financial of non-financial interests to declare.

Ethics approval

This research was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of the Faculty of Psychology and Educational Sciences of Ghent University (application number 2016/82).

Consent to participate

Written informed consent was obtained from the participants (i.e., first-year university students).

Consent for publication

Not applicable.

Availability of data and materials

The data supporting the findings of this study are protected and are not available due to data privacy laws. However, the data are available from the authors upon reasonable request and with permission of the university's Department of Educational Affairs. Variance-covariance matrices are available at https://osf.io/9bzmq/?view_only=aa49ce93049241b1a4facbc2b2088c00. The GULS items can be found in *Appendix D*, Items GULS. The reported study was not preregistered.

Code availability

Analysis codes are available at https://osf.io/9bzmq/?view_only=aa49ce93049241b1a4facbc2b2088c00

Authors' contributions

Mona Bassleer: Conceptualization, Methodology, Validation, Formal Analysis, Data Curation, Writing – Original Draft, Review & Editing, Visualization. **Stijn Schelfhout:** Conceptualization, Methodology, Validation, Formal Analysis, Writing – Review & Editing. **Lot Fonteyne:** Conceptualization, Methodology, Validation, Investigation, Data Curation, Writing – Review & Editing. **Mit Leuridan:** Conceptualization, Methodology, Validation, Writing - Review & Editing. **Wouter Duyck:** Conceptualization, Funding Acquisition, Methodology, Writing – Review & Editing, Supervision. **Nicolas Dirix:** Conceptualization, Methodology, Formal Analysis, Writing – Review & Editing, Supervision.

References

Abunawas, M. (2014). *A meta-analytic investigation of the predictive validity of the Test of English as a Foreign Language (TOEFL) scores on GPA*. https://oaktrust.library.tamu.edu/bitstream/handle/1969.1/154156/ABUNAWAS-DISSERTATION-2014.pdf?sequence=1

Baker, F. B. (2001). *The Basics of Item Response Theory. Second Edition.* MD: ERIC Clearinghouse on Assessment and Evaluation.

Berns, M., & Matsuda, K. (2006). Applied Linguistics: Overview and History. In *Encyclopedia of Language & Linguistics* (pp. 394–405). Elsevier. https://doi.org/10.1016/B0-08-044854-2/00599-X

Christensen, A. P., Garrido, L. E., & Golino, H. (2023). Unique Variable Analysis: A Network Psychometrics Method to Detect Local Dependence. *Multivariate Behavioral Research*, *58*(6), 1165–1182. https://doi.org/10.1080/00273171.2023.2194606

Clinton-Lisell, V., Taylor, T., Carlson, S. E., Davison, M. L., & Seipel, B. (2022). Performance on Reading Comprehension Assessments and College Achievement: A Meta-Analysis. *Journal of College Reading and Learning*, *52*(3), 191–211. https://doi.org/10.1080/10790195.2022.2062626

De Wachter, L., Heeren, J., Marx, S., & Huyghe, S. (2013). “Taal: een noodzakelijke, maar niet de enige voorwaarde tot studiesucces: De correlatie tussen de resultaten van een taalvaardigheidstoets en de slaagcijfers bij eerstejaarsstudenten aan de KU Leuven.” [Language Proficiency: A Necessary, but Not the Only, Condition for Study Success: A Correlation between the Results of a Language Proficiency Test and Academic

Achievement of First-Year Students.]. *Levende Talen Tijdschrift*, *14*(4), 28–36. https://lt-tijdschriften.nl/ojs/index.php/ltt/article/view/549

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. https://doi.org/10.1111/bjop.12046

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (2024). *Ethnologue: Languages of the World. Twenty-seventh edition.* Texas: SIL International. https://www.ethnologue.com/language/nld/

Elder, C. (2017). Language Assessment in Higher Education. In *Language Testing and Assessment* (pp. 271–286). Springer International Publishing. https://doi.org/10.1007/978-3-319-02261-1_35

Espinoza, O. (2007). Solving the equity–equality conceptual dilemma: a new model for analysis of the educational process. *Educational Research*, *49*(4), 343–363. https://doi.org/10.1080/00131880701717198

European Commission, D.-G. for E. Y. S. and C. (2015). *ECTS users' guide 2015*. https://data.europa.eu/doi/10.2766/87192

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, A. (2009). *COTAN beoordelingssysteem voor de kwaliteit van tests (geheel herziene versie)*. http://www.psynip.nl/website/wat-doethet- nip/tests/beoordelingsprocedure/beoordelingsprocedure

Fonteyne, L. (2017). *Constructing SIMON: a tool for evaluating personal interests and capacities to choose a post-secondary major that maximally suits the potential*. Ghent University.

Fonteyne, L., De Fruyt, F., Dewulf, N., Duyck, W., Erauw, K., Goeminne, K., Lammertyn, J., Marchant, T., Moerkerke, B., Oosterlinck, T., & Rosseel, Y. (2015). Basic mathematics test predicts statistics achievement and overall first year academic success. *European*

Journal of Psychology of Education, *30*(1), 95–118. https://doi.org/10.1007/s10212-014-0230-9

Fonteyne, L., Duyck, W., & De Fruyt, F. (2017). Program-specific prediction of academic achievement on the basis of cognitive and non-cognitive factors. *Learning and Individual Differences*, *56*, 34–48. https://doi.org/10.1016/j.lindif.2017.05.003

Gagen, T. (2019). *The predictive validity of IELTS scores: A meta-analysis*. https://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=8762&context=etd

Graham, S. (2015). Inaugural editorial for the Journal of Educational Psychology. *Journal of Educational Psychology*, *107*(1), 1–2. https://doi.org/10.1037/edu0000007

Hambleton, R., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. CA: Sage Publications.

Harris, K. R. (2003). Editorial: Is the work as good as it could be? *Journal of Educational Psychology*, *95*(3), 451–452. https://doi.org/10.1037/0022-0663.95.3.451

Hauspie, C., Schelfhout, S., Dirix, N., Fonteyne, L., Janse, M., Szmalec, A., Vereeck, A., & Duyck, W. (2024). Does Studying Latin in Secondary Education Predict Study Achievement in Academic Higher Education? *Language Learning*. https://doi.org/10.1111/lang.12639

Heeren, J., Speelman, D., & De Wachter, L. (2021). A practical academic reading and vocabulary screening test as a predictor of achievement in first-year university students: implications for test purpose and use. *International Journal of Bilingual Education and Bilingualism*, *24*(10), 1458–1473. https://doi.org/10.1080/13670050.2019.1709411

Hulstijn, J. H. (2015). *Language Proficiency in Native and Non-native Speakers*. John Benjamins B.V.

Ihlenfeldt, S. D., & Rios, J. A. (2023). A meta-analysis on the predictive validity of English language proficiency assessments for college admissions. *Language Testing*, *40*(2), 276–299. https://doi.org/10.1177/02655322221112364

Jansen, C., De Wachter, L., Van Dun, P., & Frik, T. (2022). *Taalcompetentie in het Nederlands van Nederlandstalige studenten in het hoger onderwijs in Nederland en Vlaanderen*. https://taalunie.org/publicaties/213/taalcompetentie-in-het-nederlands-van-nederlandstalige-studenten-in-het-hoger-onderwijs-in-nederland-en-vlaanderen

Katz, D., Clairmont, A., & Wilton, M. (2021). *Measuring what Matters: Introduction to Rasch Analysis in R*. https://bookdown.org/dkatz/Rasch_Biome/

Kennedy, T. J., & Sundberg, C. W. (2020). *21st Century Skills* (pp. 479–496). https://doi.org/10.1007/978-3-030-43620-9_32

Kintsch, W. (2013). Revisiting the Construction–Integration Model of Text Comprehension and Its Implications for Instruction. In *Theoretical Models and Processes of Reading* (pp. 807–839). International Reading Association. https://doi.org/10.1598/0710.32

Knoch, U., & Elder, C. (2013). A framework for validating post-entry language assessments (PELAs). *Papers in Language Testing and Assessment*, *2*(2), 48–66. https://arts.unimelb.edu.au/__data/assets/pdf_file/0007/1771306/4_Knoch_Elder_0.pdf

Krehbiel, T. C. (2004). Correlation Coefficient Rule of Thumb. *Decision Sciences Journal of Innovative Education*, *2*(1), 97–100. https://doi.org/10.1111/j.0011-7315.2004.00025.x

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Marcoulides, K. M., & Raykov, T. (2019). Evaluation of Variance Inflation Factors in Regression Models Using Latent Variable Modeling Methods. *Educational and*

Psychological Measurement, *79*(5), 874–882. https://doi.org/10.1177/0013164418817803

Meeter, M. (2023). Predicting retention in higher education from high-stakes exams or school GPA. *Educational Assessment, 28*, 1-10. https://doi.org/10.1080/10627197.2022.2130748

Min, S., & Aryadoust, V. (2021). A systematic review of item response theory in language assessment: Implications for the dimensionality of language ability. *Studies in Educational Evaluation*, *68*, 100963. https://doi.org/10.1016/j.stueduc.2020.100963

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory (3rd ed.)*. New York: McGraw-Hill.

OECD. (2019). *An OECD Learning Framework 2030* (pp. 23–35). https://doi.org/10.1007/978-3-030-26068-2_3

OECD. (2023). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. https://doi.org/https://doi.org/10.1787/53f23881-en

Pyle, N., Vasquez, A. C., Lignugaris/Kraft, B., Gillam, S. L., Reutzel, D. R., Olszewski, A., Segura, H., Hartzheim, D., Laing, W., & Pyle, D. (2017). Effects of Expository Text Structure Interventions on Comprehension: A Meta-Analysis. *Reading Research Quarterly, 52*(4), 469–501. https://doi.org/10.1002/rrq.179

Read, J. (2015). *Assessing English Proficiency for University Study*. Palgrave Macmillan UK. https://doi.org/10.1057/9781137315694

Read, J. (2016). *Some Key Issues in Post-Admission Language Assessment* (pp. 3–20). https://doi.org/10.1007/978-3-319-39192-2_1

Reise, S., Bonifay, W., & Haviland, M. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*(2), 129-140, https://doi.org/10.1080/00223891.2012.725437

Reise, S. Scheines, R., Widaman, K., & Haviland, M. (2023). Multidimensionality and Structural Coefficient Bias in Structural Equation Modeling A Bifactor Perspective. *Educational and Psychological Measurement, 73*(1), 5-26. https://doi.org/10.1177/0013164412449831

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353–387. https://doi.org/10.1037/a0026838

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

Schelfhout, S. (2019). *The Influence of Study Interests and (Non-)Cognitive Predictors on Study Choice and Study Achievement in Flemish Higher Education* [Dissertation]. Ghent University.

Siegelman, N., Schroeder, S., Acartürk, C., Ahn, H.-D., Alexeeva, S., Amenta, S., Bertram, R., Bonandrini, R., Brysbaert, M., Chernova, D., Da Fonseca, S. M., Dirix, N., Duyck, W., Fella, A., Frost, R., Gattei, C. A., Kalaitzi, A., Kwon, N., Lõo, K., … Kuperman, V. (2022). Expanding horizons of cross-linguistic research on reading: The Multilingual Eye-movement Corpus (MECO). *Behavior Research Methods*, *54*(6), 2843–2863. https://doi.org/10.3758/s13428-021-01772-6

Sohn, H., Acosta, K., Brownell, M. T., Gage, N. A., Tompson, E., & Pudvah, C. (2023). A Meta-Analysis of Interventions to Improve Reading Comprehension Outcomes for Adolescents with Reading Difficulties. *Learning Disabilities Research & Practice, 38*(2), 85–103. https://doi.org/10.1111/ldrp.12307

Stricker, L. J. (2004). The performance of native speakers of English and ESL speakers on the computer-based TOEFL and GRE General Test. *Language Testing*, *21*(2), 146–173. https://doi.org/10.1191/0265532204lt279oa

Tavakol, M., & Wetzel, A. (2020). Factor Analysis: a means for theory and instrument development in support of construct validity. *International Journal of Medical Education*, *11*, 245–247. https://doi.org/10.5116/ijme.5f96.0f4a

Ünal, Z. E., Greene, N. R., Lin, X., & Geary, D. C. (2023). What Is the Source of the Correlation Between Reading and Mathematics Achievement? Two Meta-analytic Studies. *Educational Psychology Review*, *35*(1), 4. https://doi.org/10.1007/s10648-023-09717-5

van Dijk, T. (2015). Tried and tested. *Tijdschrift Voor Taalbeheersing*, *37*(2), 159–186. https://doi.org/10.5117/TVT2015.2.VAND

Westrick, P.A., Le, H., Robbins, S.B., Radunzel, J.M.R. & Schmidt, F.L. (2015). College Performance and Retention: A Meta-Analysis of the Predictive Validities of ACT ® Scores, High School Grades, and SES. *Educational Assessment, 20*, 23-45. https://doi.org/10.1080/10627197.2015.997614

Wingate, U. (2015). *Academic Literacy and Student Diversity*. Multilingual Matters. https://doi.org/10.21832/9781783093496

Wongtrirat, R. (2010). *English language proficiency and academic achievement of international students: A meta-analysis*. https://digitalcommons.odu.edu/cgi/viewcontent.cgi?article=1183&context=efl_etds

York, T. Y., Gibson, C., & Rankin, S. (2015). Defining and Measuring Academic Success. *Practical Assessment, Research & Evaluation*, *20*(5), 1–20. http://pareonline.net/getvn.asp?v=20&n=5

Appendix A: Extended Sample Size Data and Results

Table A1

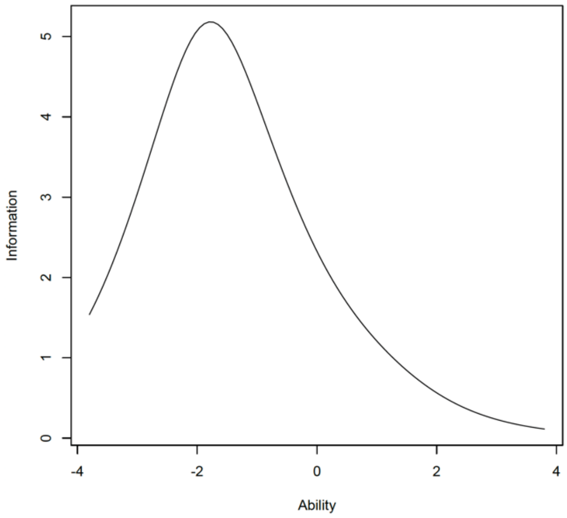
Subsample Sizes per Study Program

	Sex		SES		Total
	Male	Female	Low	High	
Basic mathematics					
Applied Language Studies	105	431	138	398	536
Communication Sciences	110	361	81	390	471
Criminological Sciences	243	928	345	826	1,171
Educational Sciences	33	604	154	483	637
History	242	166	95	313	408
Law	579	1,441	547	1,473	2,020
Pharmaceutical Sciences	220	921	290	851	1,141
Physical Therapy and Motor Rehabilitation	648	1,150	338	1,460	1,798
Political Sciences	258	207	121	344	465
Psychology	395	2,292	769	1,918	2,687
Advanced mathematics					
Biochemistry and Biotechnology	180	250	92	338	430
Biomedical Sciences	235	767	260	742	1,002
Bioscience Engineering	589	601	179	1,011	1,190
Business Administration	938	831	401	1,368	1,769
(Applied/Business) Economics	1,191	761	350	1,602	1,952
Engineering Technology	1,290	168	288	1,170	1,458
Total	7,256	11,879	4,448	14,687	19,135

Note. The advanced mathematics study programs are associated with more advanced mathematical curricula. Through the SIMON project (Fonteyne, 2017), students enrolling in basic mathematics study programs complete a less difficult math assessment compared to students enrolling in more advanced mathematics study programs. For more information on the predictive validity of this basic mathematics test, see Fonteyne and colleagues (2015).

Figure A1

Test Information Curve Three-Parameter Model Sample 2017-2018 to 2019-2020

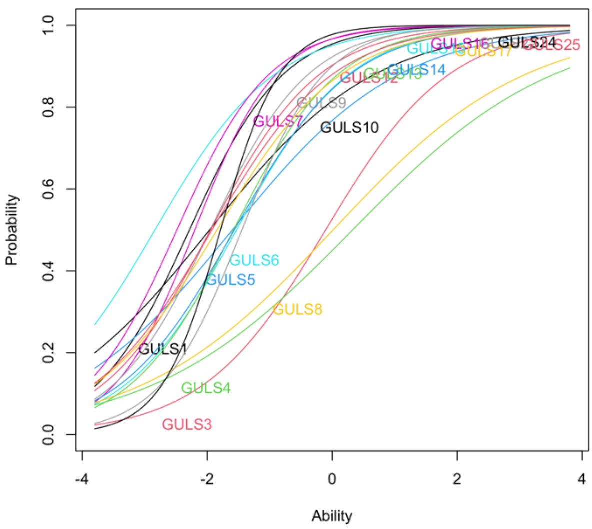


Note. N = 12,527. The three-parameter model offers 82% test information within the lower range of language proficiency ability [-6, 0].

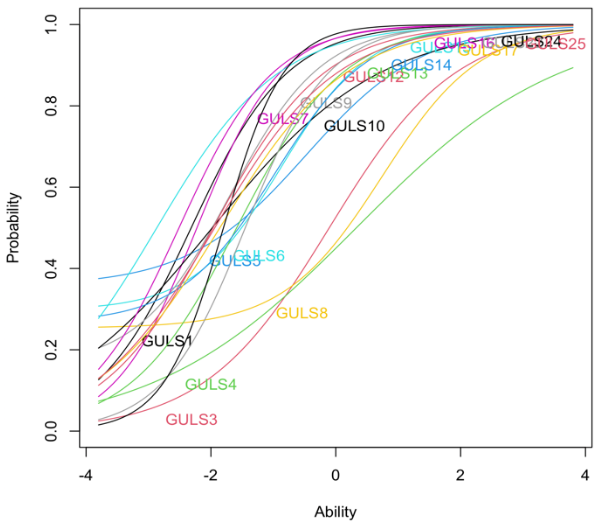
Figure A2

Item Characteristics Curves Two-and Three-Parameter Model Sample 2017-2018 to 2019-2020

Two-Parameter Model



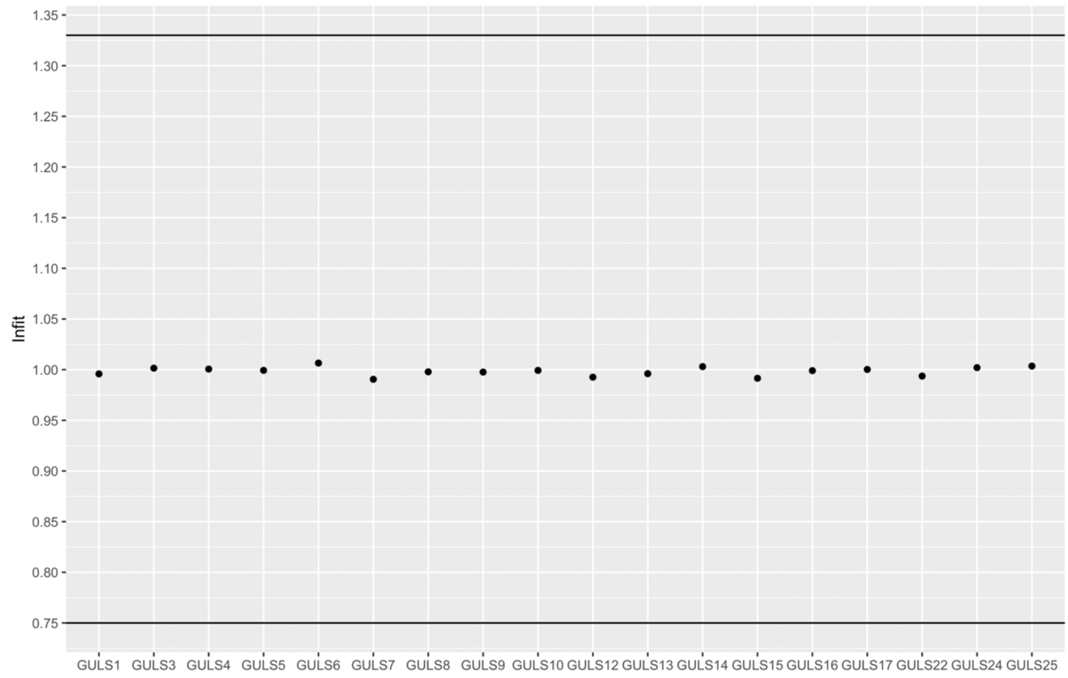
Three-Parameter Model

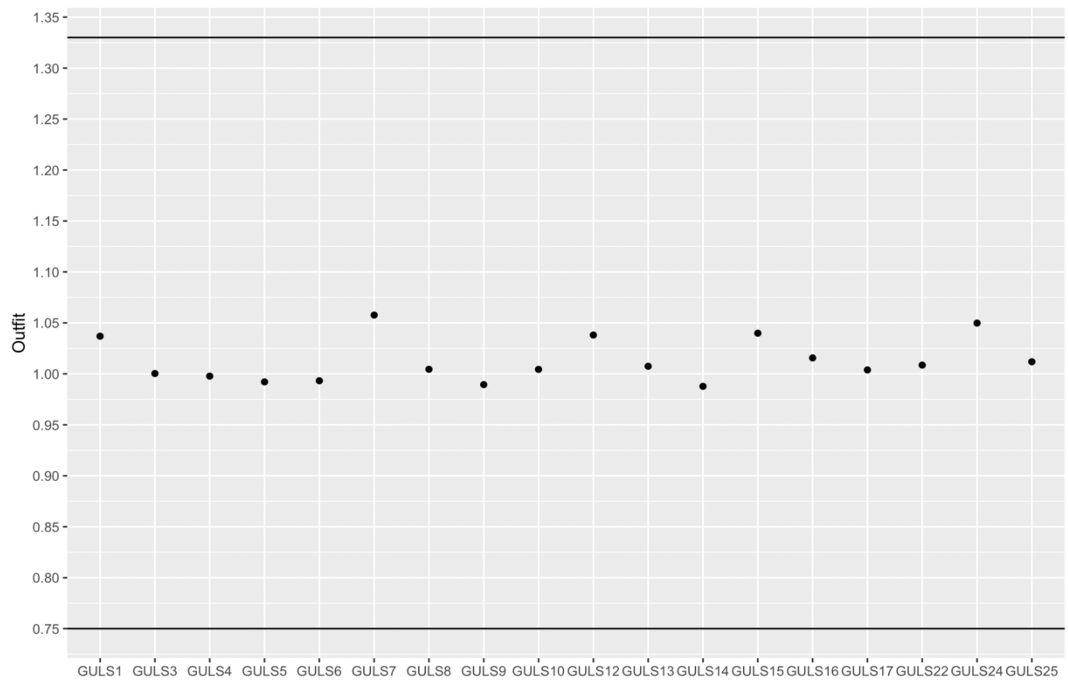


Note. N = 12,527. Item difficulties range from -2.84 to 0.33, and item discriminations vary between 0.62 and 2.11.

Figure A3

Item Infits (Top) and Outfits (Down) Sample 2017-2018 to 2019-2020

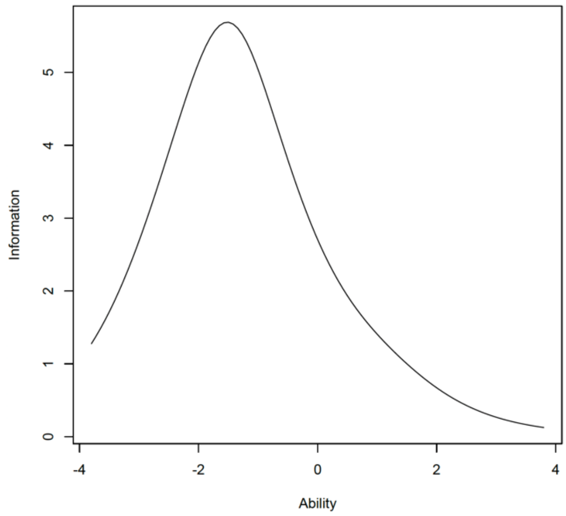




Note. N = 12,527. All item in-and outfit statistics meet the desired criteria of 0.75 to 1.33 (Katz et al., 2021).

Figure A4

Test Information Curve Three-Parameter Model Sample 2020-2021 to 2022-2023

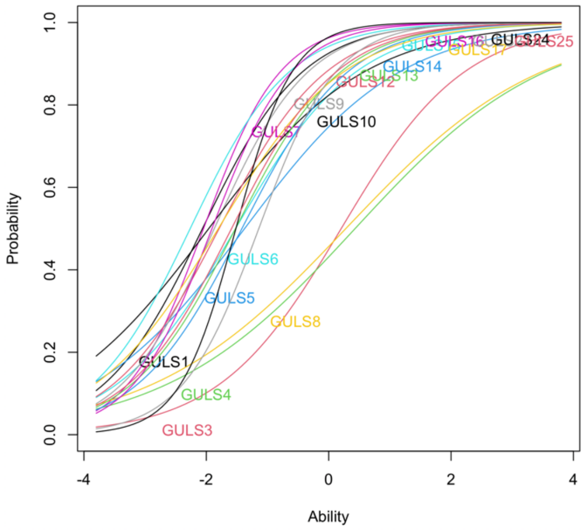


Note. N = 17,204. The three-parameter model offers 81% test information within the lower range of language proficiency ability [-6, 0].

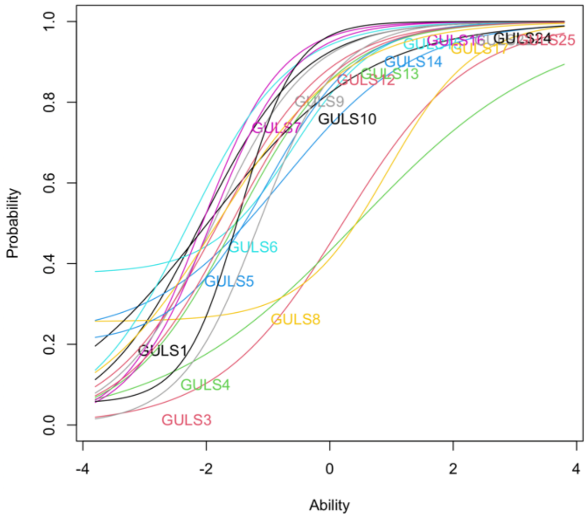
Figure A5

Item Characteristics Curves Two-and Three-Parameter Model Sample 2020-2021 to 2022-2023

Two-Parameter Model



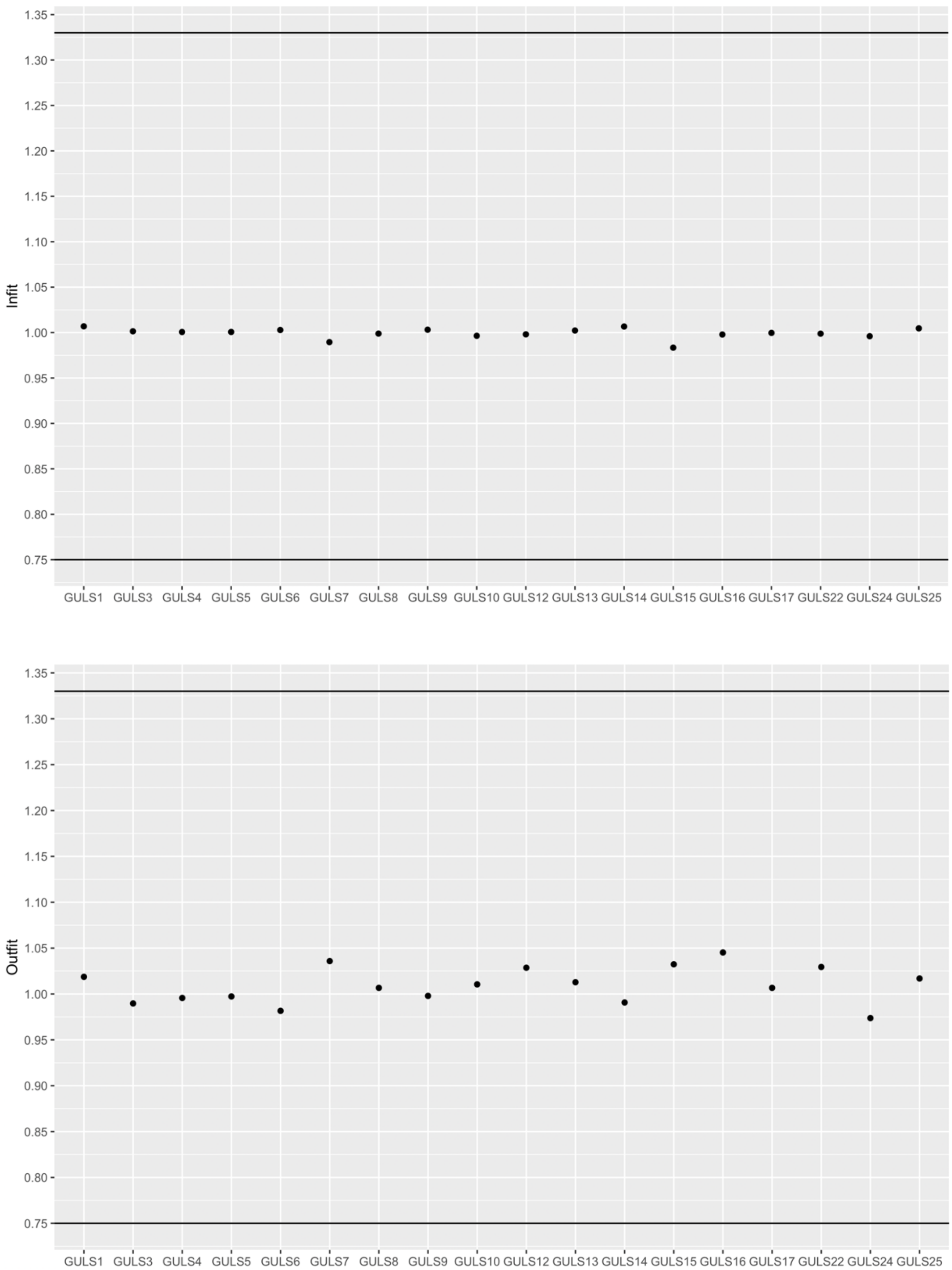
Three-Parameter Model



Note. N = 17,204. Item difficulties span from -2.26 to 0.43, while item discriminations range between 0.62 and 2.17.

Figure A6

Item Infits (Top) and Outfits (Down) Sample 2020-2021 to 2022-2023

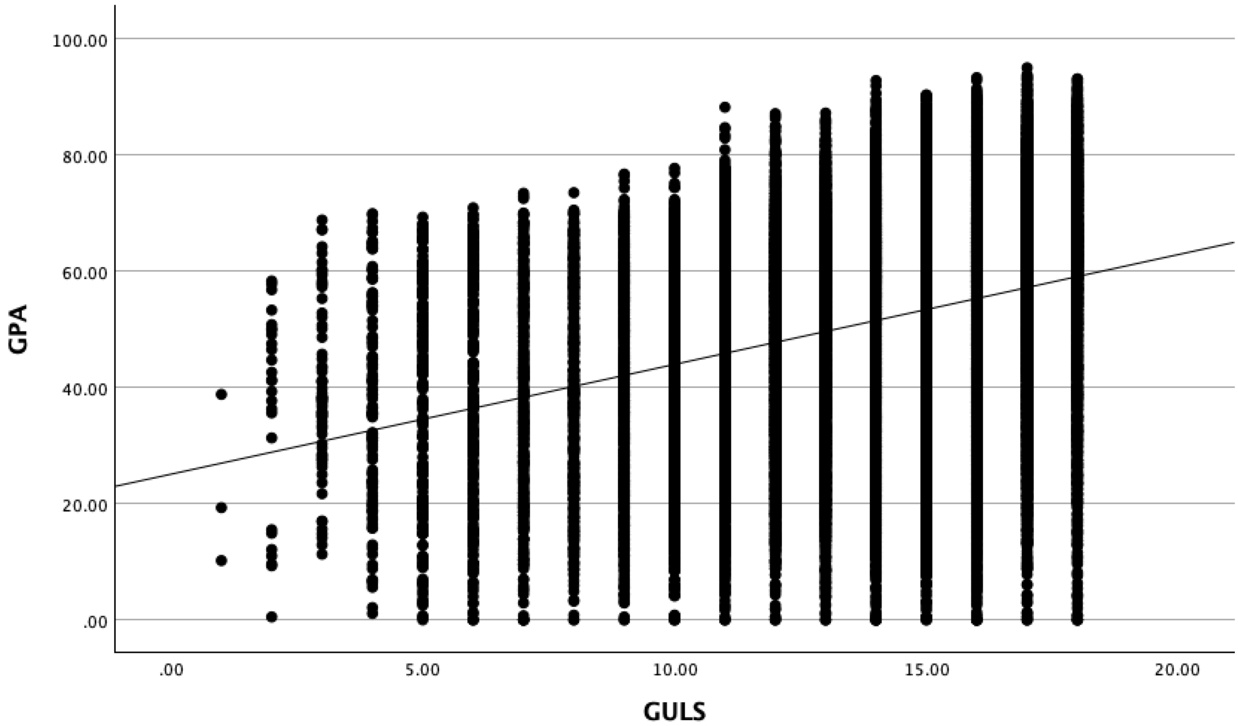


Note. N = 17,204. All item in-and outfit statistics meet the desired criteria of 0.75 to 1.33 (Katz et al., 2021).

Figure A7

Scatter Plots GULS and GPA (A) and Study Success (B)

(A)



(B)

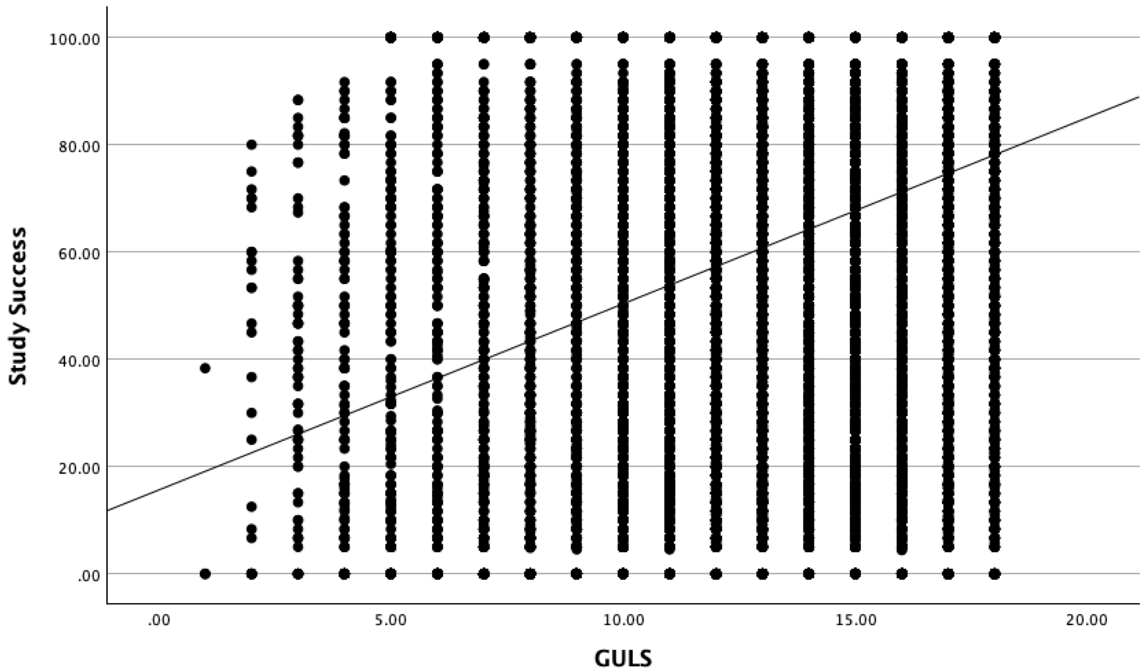


Table A3

Correlation Matrix Across and per Subsample

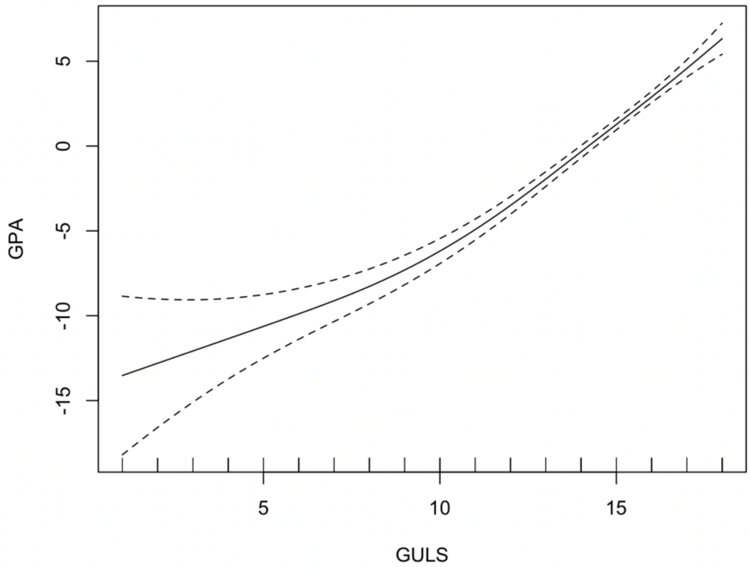
	GPA	Study Success	GULS	Math Hours SE	Vocabulary	SES[a]	Sex[b]	Math Category[c]
GPA	1	.94**	.27**	.25**	.15**	.13**	.08**	.05**
Subsample 1	1	.94**	.24**	.19**	.14**	.09**	.12**	.02
Subsample 2	1	.94**	.28**	.29**	.16**	.17**	.04**	.07**
Study Success	.94**	1	.24**	.23**	.14**	.13**	.06**	.06**
Subsample 1	.94**	1	.22**	.17**	.13**	.09**	.11**	.02
Subsample 2	.94**	1	.26**	.27**	.15**	.17**	.03**	.08**
GULS	.27**	.24**	1	.09**	.32**	.08**	.02*	-.03**
Subsample 1	.24**	.22**	1	.09**	.30**	.06**	-.03**	.00
Subsample 2	.28**	.26**	1	.09**	.33**	.10**	.05**	-.06**
Math Hours SE	.25**	.23**	.09**	1	.05**	.11**	-.18**	.42**
Subsample 1	.19**	.17**	.09**	1	.05**	.11**	-.21**	.43**
Subsample 2	.29**	.27**	.09**	1	.05**	.11**	-.16**	.41**
Vocabulary	.15**	.14**	.32**	.05**	1	.09**	.02*	-.02*
Subsample 1	.14**	.13**	.30**	.05**	1	.09**	-.003	-.002
Subsample 2	.16**	.15**	.33**	.05**	1	.09**	.03**	-.03**

Note. *$p < .05$, **$p < .01$. Subsamples 1 ($N = 8,244$) and 2 ($N = 10,891$) respectively encompass students from academic years 2017-2018 to 2019-2020 and 2020-2021 to 2022-2023. [a]Low SES = 0, High SES = 1; [b]Male = 0, Female = 1; [c]Basic mathematics study programs = 0; Advanced mathematics study programs = 1. SE = Secondary Education. The advanced mathematics study programs are associated with more advanced mathematical curricula. Through the SIMON project (Fonteyne, 2017), students enrolling in basic mathematics study programs complete a less difficult math assessment compared to students enrolling in more advanced mathematics study programs. Pearson correlation coefficients are shown for two continuous variables; [a,b,c]Point-biserial correlation coefficients are shown. The advanced mathematics study programs are associated with more advanced mathematical curricula.

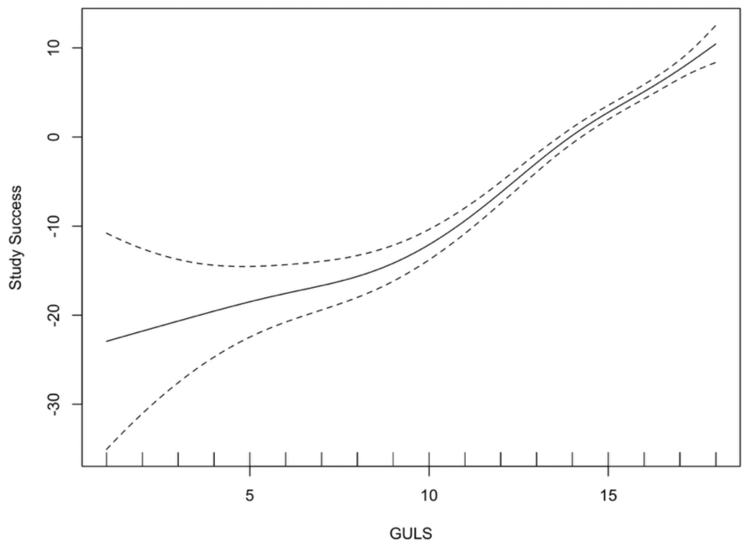
Figure A8

Plots Generalized Additive Models (GAMs) GULS and GPA (A) and Study Success (B)

(A)



(B)



Note. The y-axis shows the estimated smooth effect of GULS on GPA (A) and study success (B). A value of 0 indicates that the predicted outcome equals the overall sample mean. Positive values indicate an increase, negative values a decrease, relative to the mean.

Table A4

GAM-Based Estimated Smooth Effects and Confidence Intervals of GULS on GPA

GULS	Estimated smooth	95% *CI*	
		Lower	Upper
1	-13.52	-22.68	-4.36
2	-12.78	-20.15	-5.41
3	-12.03	-17.86	-6.21
4	-11.41	-16.16	-6.66
5	-10.66	-14.38	-6.95
6	-9.90	-12.85	-6.96
7	-9.11	-11.51	-6.71
8	-8.24	-10.23	-6.24
9	-7.23	-8.92	-5.55
10	-6.28	-7.74	-4.82
11	-4.98	-6.20	-3.76
12	-3.51	-4.53	-2.49
13	-1.92	-2.76	-1.07
14	-0.26	-0.97	0.46
15	1.40	0.78	2.02
16	2.79	2.14	3.44
17	4.52	3.54	5.51
18	5.59	4.53	8.12

Table A5

GAM-Based Estimated Smooth Effects and Confidence Intervals of GULS on Study Success

GULS	Estimated smooth	95% *CI*	
		Lower	Upper
1	-22.94	-47.21	1.34
2	-21.76	-40.11	-3.41
3	-20.60	-34.18	-7.02
4	-19.64	-30.19	-9.09
5	-18.55	-26.58	-10.52
6	-17.58	-24.04	-11.12
7	-16.68	-22.13	-11.23
8	-15.61	-20.28	-10.93
9	-14.04	-18.03	-10.06
10	-12.25	-15.71	-8.80
11	-9.52	-12.41	-6.62
12	-6.27	-8.71	-3.84
13	-2.84	-4.90	-0.78
14	-0.32	-1.43	2.08
15	2.96	1.40	4.53
16	4.95	3.36	6.55
17	7.53	5.46	9.60
18	10.45	6.29	14.62

References

Fonteyne, L., De Fruyt, F., Dewulf, N., Duyck, W., Erauw, K., Goeminne, K., Lammertyn, J., Marchant, T., Moerkerke, B., Oosterlinck, T., & Rosseel, Y. (2015). Basic mathematics test predicts statistics achievement and overall first year academic success. *European Journal of Psychology of Education*, *30*(1), 95–118. https://doi.org/10.1007/s10212-014-0230-9

Fonteyne, L., Duyck, W., & De Fruyt, F. (2017). Program-specific prediction of academic achievement on the basis of cognitive and non-cognitive factors. *Learning and Individual Differences*, *56*, 34–48. https://doi.org/10.1016/j.lindif.2017.05.003

Katz, D., Clairmont, A., & Wilton, M. (2021). *Measuring what Matters: Introduction to Rasch Analysis in R*. https://bookdown.org/dkatz/Rasch_Biome/

Appendix B: Results Study Success After the Second-Chance Exam Period

Predictive Validity

For the analyses across study programs and study program-specific, no VIF value exceeds 2, indicating the absence of multicollinearity issues.

Across Study Programs

Subsample 1. For the 2017-2018 to 2019-2020 data across 16 study programs, the correlation matrix can be found in *Appendix A*, Table A3. GULS (M = 14.1, SD = 2.9) has a significant effect on study success after the second-chance exam period (hereafter: study success), when controlled for sex, SES, mathematics hours secondary education, vocabulary knowledge, and mathematics category ($t(8{,}237) = 16.55$, $p < .001$, $B = 2.16$, $\beta = 0.18$). Individually, the explained variance of GULS in study success is $R^2 = .05$. GULS's unique contribution to the explained variance in study success is $R^2 = .03$. The difference in explained variance between the full model ($R^2 = .10$) and the model without GULS is significant ($F_{Change}(1, 8{,}237) = 273.80$, $p < .001$). The internal reliability $\omega = .73$.

Subsample 2. For the 2020-2021 to 2022-2023 data across 16 study programs, the correlation matrix is available in *Appendix A*, Table A3. GULS (M = 13.6, SD = 3.1) has a significant effect on study success, when controlled for sex, SES, mathematics hours secondary education, vocabulary knowledge, and mathematics category ($t(10{,}884) = 21.28$, $p < .001$, $B = 2.37$, $\beta = 0.20$). Individually, the explained variance of GULS in study success is $R^2 = .07$. GULS's unique contribution to the explained variance in study success is $R^2 = .04$. The difference in explained variance between the full model ($R^2 = .15$) and the model without GULS is significant ($F_{Change}(1, 10{,}884) = 452.79$, $p < .001$). The internal reliability $\omega = .75$.

Study Program-Specific

The data from 2017-2018 to 2022-2023 shows that GULS significantly predicts study success for all the basic mathematics study programs (M range [13.2, 15.2], SD range [2.2, 3.1]), when controlled for sex, SES, mathematics hours secondary education, vocabulary knowledge and mathematical proficiency (*p*'s < .001). Individually, the explained variance of GULS in study success is R^2 = .05 (Physical Therapy and Motor Rehabilitation, and Political Sciences) to R^2 = .18 (Applied Language Studies). GULS's unique contribution to the explained variance in study success varies from R^2 = .01 (Physical Therapy and Motor Rehabilitation) to R^2 = .10 (Applied Language Studies). The differences in explained variance between the full models and the models without GULS are significant (*p*'s < .001).

Across the advanced mathematics study programs (*M* range [12.7, 14.9], *SD* range [2.7, 3.3], the study programs Biochemistry and Biotechnology, and Engineering Technology diverge from the observed pattern where GULS significantly predicts study success, when controlled for sex, SES, mathematics hours secondary education, vocabulary knowledge and mathematical proficiency (*p* range [< .001, .657]). Individually, the explained variance of GULS in study success is R^2 = .02 (Biochemistry and Biotechnology, and Engineering Technology) to R^2 = .06 (Business Administration, and (Applied/Business) Economics). GULS's unique contribution to the explained variance in study success varies from R^2 < .01 (Biochemistry and Biotechnology, and Engineering Technology) to R^2 = .04 (Business Administration). The differences in explained variance between the full models and the models without GULS are significant, except for the study programs Biochemistry and Biotechnology, and Engineering Technology (*p* range [< .001, .657]). For details on GULS's descriptives, predictive validity measures for study success, reliability, and the total explained variance of the full model per study program, we refer to Table B1.

Table B1

GULS's Descriptives, Predictive Validity Measures for Study Success, Reliability, and Total Explained Variance Full Model per Study Program

	M (*SD*)	*t*	*B* (*b*)	R^2			*ω*
				Individual	Unique	Total	
Basic mathematics							
Applied Language Studies	14.7 (2.6)	8.38***	4.83 (0.58)	.18	.10***	.24	.70
Communication Sciences	14.2 (2.6)	4.07***	2.65 (0.65)	.07	.03***	.12	.68
Criminological Sciences	13.5 (2.8)	6.55***	2.65 (0.41)	.07	.03***	.13	.68
Educational Sciences	13.9 (2.6)	5.89***	2.11 (0.36)	.08	.05***	.13	.63
History	15.2 (2.2)	4.08***	3.56 (0.87)	.11	.04***	.15	.64
Law	14.3 (2.7)	12.16***	3.82 (0.31)	.14	.06***	.20	.72
Pharmaceutical Sciences	13.7 (3.1)	5.91***	1.89 (0.32)	.10	.03***	.17	.77
Physical Therapy and Motor Rehabilitation	13.2 (3.1)	5.31***	1.23 (0.23)	.05	.01***	.18	.74
Political Sciences	14.5 (2.5)	3.95***	2.81 (0.71)	.05	.03***	.06	.68
Psychology	13.8 (2.8)	11.33***	2.81 (0.25)	.09	.04***	.18	.69
Advanced mathematics							
Biochemistry and Biotechnology	14.1 (3.1)	1.05	0.64 (0.61)	.02	< .01	.10	.78
Biomedical Sciences	13.7 (3.1)	3.65***	1.45 (0.40)	.05	.01***	.12	.76
Bioscience Engineering	14.9 (2.7)	2.54*	0.97 (0.38)	.05	< .01*	.19	.78
Business Administration	12.7 (3.3)	9.36***	2.20 (0.24)	.06	.04***	.14	.75
(Applied/Business) Economics	13.9 (3.1)	6.75***	1.63 (0.24)	.06	.02***	.13	.78
Engineering Technology	13.4 (3.2)	0.44	0.13 (0.30)	.02	< .01	.13	.77

Note. *$p < .05$, **$p < .01$, ***$p < .001$. *ω* = McDonald's *ω* as reliability measure. R^2 = Adjusted R-squared. Total R^2 represents the explained variance by the full model, whereas unique R^2 quantifies the specific contribution of GULS to study success, distinct from other predictors (i.e., sex, SES, mathematics hours secondary education, vocabulary knowledge, and mathematical proficiency). Meanwhile, individual R^2 reflects the overall contribution of GULS to study success, regardless of additional predictors. The advanced mathematics study programs are associated with more advanced mathematical curricula. Through the SIMON project (Fonteyne, 2017), students enrolling in basic mathematics study programs complete a less difficult math assessment compared to students enrolling in more advanced mathematics study programs.

References

Fonteyne, L. (2017). *Constructing SIMON: a tool for evaluating personal interests and capacities to choose a post-secondary major that maximally suits the potential*. Ghent University.

Appendix C: Preliminary Analyses Pilot GULS

Participants

Between September 20 and October 6, 2016, $N = 972$ anonymous secondary school students, primarily from the East Flanders region, completed the online GULS with 45 items: $n = 726$ from ASO (general secondary education), $n = 216$ from TSO (technical secondary education), and $n = 30$ from BSO (vocational secondary education).

General Findings

Table C1 presents the mean total scores and standard deviations across the different educational tracks. The variance between these tracks is significant ($F(2,971) = 181.34$, $p < .001$, $R^2 = .27$). Given that the test is not specifically tailored for any track, we will further uniformly analyze GULS.

Table C1

Descriptives Across and per Educational Track

	M	*SD*	*n*
ASO	33.70	5.56	726
TSO	26.05	6.44	216
BSO	22.57	7.23	30
Total	31.66	6.82	972

Note. ASO = general secondary education, TSO = technical secondary education, BSO = vocational secondary education.

Item Analysis

The test is divided into eight categories: Text Structure Recognition (TR - 5 items), Word Meaning (WM - 6 items), Paragraph Construction (PC - 4 items), Text Comprehension (TC - 4 items), Contextual Word Filling (CWF - 6 items), Signal Words (SW - 5 items), Correct Form (CF - 10 items), and Word Combinations (WC - 5 items). Table C2 shows the scoring percentages for each item in descending order.

Table C2

Item Scoring Percentages in Descending Order

Item	*M*	*SD*
WM4	.97	.18
WM2	.96	.19
CF2	.94	.23
WM3	.94	.23
CWF1	.93	.25
CWF5	.92	.27
WM1	.91	.29
WM6	.87	.34
TC4	.87	.34
CF9	.86	.35
WC2	.86	.35
CWF2	.85	.35
CF4	.85	.36
TR2	.84	.37
PC4	.83	.38
CF1	.82	.39
CF3	.81	.39
WC3	.80	.40
TR4	.77	.42
CWF3	.77	.42
SW1	.75	.43
SW5	.75	.43
SW2	.75	.43
WC4	.73	.45
CF8	.72	.45
SW3	.72	.45
PC3	.71	.45
TR1	.70	.46
WC5	.68	.47
TC2	.66	.47
CWF4	.66	.47
SW4	.63	.48
TR5	.63	.48
TR3	.58	.50
PC1	.56	.50
CF5	.53	.50
TC1	.53	.50
PC2	.53	.50
WM5	.43	.50
CF6	.41	.49
CF10	.39	.49
WC1	.37	.48
CWF6	.36	.48
TC3	.34	.47
CF7	.16	.37

Note. *N* = 972. All the items are dichotomous. TR = Text Structure Recognition, WM = Word Meaning, PC = Paragraph Construction, TC = Text Comprehension, CWF = Contextual Word Filling, SW = Signal Words, CF = Correct Form, WC = Word Combinations.

Psychometrics

After analysis, an IRT model on dichotomous data was performed. This model represents the probability of a correct answer given the underlying latent language proficiency of the participant, summarized by the formula:

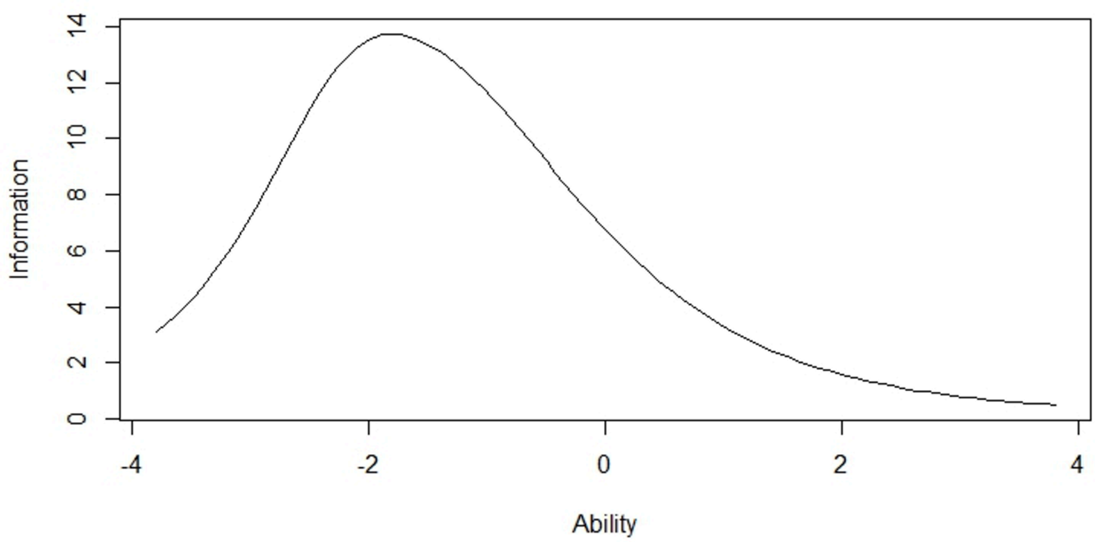
$$P\ (x_{im} = 1 \mid z_m) = g\{\alpha_i(z_m - \beta_i)\}$$

where x_{im} is the score of participant m on item *i*, z_m is the latent proficiency of participant *m,* α_i is the discrimination parameter for item *i* and β_i is the difficulty for item *i*. The logit (or probit) link function is represented by *g*. Implementing a guessing parameter c_i was also explored, which provided a better fit ($p < .001$). Strictly speaking, however, this model could not be applied because some items scored 0 due to blank answers. As the models were similar, only the model without the guessing parameter is discussed further. Nevertheless, it is advisable to consider the model with the guessing parameter in future decisions for a shorter test. The number of blanks is only 0.2% of the total answers.

GULS achieves an internal consistency of Cronbach's $\alpha = 0.85$, considered good to very good in literature (Evers et al., 2009; Nunnally & Bernstein, 1994). Correlations of individual items with the total score (excluding the individual item) range from $r = .11$ (CWF6) to $r = .48$ (PC4), supporting evidence for a common latent construct (language proficiency). Figure C1 shows the test information curve for the model without the guessing parameter. The information in the interval [-6, 0] is 79%, indicating that the test provides information on respondents with lower language proficiency, aligning with the SIMON philosophy. The test is not too difficult, but it is recommended to avoid selecting the easiest items for the final test to maintain sufficient variance. Table C3 shows the item difficulty (*b*) and discrimination (*a*) parameters.

Figure C1

Test Information Curve Two-Parameter Model



Note. The two-parameter model offers 79% test information within the lower range of language proficiency ability [-6, 0].

Table C3

Item Difficulty and Discrimination Parameters

Item	Cluster	*a*	*b*
1	TR1	1.14	-0.92
2	TR2	1.01	-1.90
3	TR3	0.66	-0.51
4	TR4	1.32	-1.23
5	TR5	0.50	-1.13
6	WM1	1.40	-2.10
7	WM2	2.33	-2.22
8	WM3	1.97	-2.12
9	WM4	2.41	-2.22
10	WM5	0.57	0.57
11	WM6	0.90	-2.40
12	PC1	0.57	-0.48
13	PC2	0.65	-0.18
14	PC3	0.71	-1.41
15	PC4	1.85	-1.30
16	TC1	0.88	-0.17
17	TC2	1.32	-0.68
18	TC3	0.70	1.04
19	TC4	1.41	-1.74
20	CWF1	1.50	-2.31
21	CWF2	1.72	-1.50
22	CWF3	1.30	-1.19
23	CWF4	1.00	-0.79
24	CWF5	1.36	-2.33
25	CWF6	0.28	2.12
26	SW1	1.31	-1.13
27	SW2	0.40	-2.82
28	SW3	1.20	-1.01
29	SW4	1.23	-0.58
30	SW5	0.78	-1.62
31	CF1	1.05	-1.72
32	CF2	1.15	-2.92
33	CF3	0.67	-2.37
34	CF4	1.59	-1.51
35	CF5	0.55	-0.25
36	CF6	0.58	0.68
37	CF7	0.42	4.08
38	CF8	1.22	-1.01
39	CF9	1.32	-1.78
40	CF10	0.41	1.10
41	WC1	1.27	0.54
42	WC2	1.40	-1.71
43	WC3	1.53	-1.26
44	WC4	1.25	-1.01
45	WC5	1.41	-0.75

Note. TR = Text Structure Recognition, WM = Word Meaning, PC = Paragraph Construction, TC = Text Comprehension, CWF = Contextual Word Filling, SW = Signal Words, CF = Correct Form, WC = Word Combinations. The item difficulty parameters (b) are classified as follows: $b < -2.00$ (very easy), $-2.00 \leq b \leq 2.00$ (moderately difficult), and $b > 2.00$ (very difficult) (Hambleton et al., 1991). The item discrimination parameters (a) are categorized in the following way: $0.01 \leq a \leq 0.34$ (very low), $0.35 \leq a \leq 0.64$ (low), $0.65 \leq a \leq 1.34$ (moderate), $1.35 \leq a \leq 1.69$ (high), $a \geq 1.70$ (very high) (Baker, 2001). The items in bold are those that were included in the GULS with 25 items (see also Conclusion and Recommendations).

Conclusion and Recommendations

GULS achieves very good internal consistency and primarily discriminates at the lower end of the latent language proficiency spectrum, in line with the SIMON philosophy. Significant differences between ASO, TSO, and BSO tracks are noted, with lower-scoring tracks exhibiting higher variance, indicating that the language test focuses on the individual rather than the track.

It is recommended not to select the easiest items and to consider the model that corrects for guessing at the item level. The item selection (25 items) to reduce GULS for further validation with first-year higher education students, should consider the item difficulties, discriminatory powers (see bold items in Table C3). We advise to exclude the items related to word meaning in this 25-item GULS, as simple vocabulary knowledge is already assessed in SIMON through the LexTALE test (Lemhöfer & Broersma, 2012).

References

Baker, F. B. (2001). *The Basics of Item Response Theory. Second Edition.* MD: ERIC Clearinghouse on Assessment and Evaluation.

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, A. (2009). *COTAN beoordelingssysteem voor* *de kwaliteit van tests (geheel herziene versie)*. http://www.psynip.nl/website/wat-doethet- nip/tests/beoordelingsprocedure/beoordelingsprocedure

Hambleton, R., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. CA: Sage Publications.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory (3rd ed.)*. New York: McGraw-Hill.

Appendix D: Items GULS

GULS1

Dat de wielersport een omgeving is waarin jargon goed gedijt, is niet verwonderlijk. Dat hebben we in de eerste plaats te danken aan de vroegste wielerverslaggevers, begin twintigste eeuw. Die hadden geen radio en tv, en dus geen geluid of beeld, om de prestaties van de renners te beschrijven. Ze zetten daarom al hun literaire vaardigheden in om de heroïek en dramatiek van de wedstrijden met de pen te beschrijven. In de tweede plaats helpt ook het wielrennen zelf, een sport van grootse tegenstellingen: soms veel tegenslag, soms een beetje geluk, urenlang lijden en kort maar hevig triomferen. Die elementen moeten commentatoren vatten in plastische begrippen zodat wielerliefhebbers de inspanning en de emotie van de renners bijna zelf ervaren. Tot slot is de wielrennerij een vrij besloten wereld: renners, entourage en journalisten leven soms heel dicht op elkaar, dagen- of zelfs wekenlang bij een rittenwedstrijd of grote ronde. In die relatief kleine wereld kan makkelijk een eigen 'taaltje' ontstaan.

(Van der Gucht, F., De Caluwe, J., van der Sijs, N. en Janssen, M. (te verschijnen), Atlas van het Nederlands in Vlaanderen [werktitel], Lannoo)

A. De auteur maakt een vergelijking.

B. De auteur onderbouwt een stelling met argumenten.

C. De auteur beschrijft een opeenvolging in de tijd.

D. De auteur stelt een probleem vast en zoekt een oplossing.

GULS2

Bilharzia, of schistosomiasis, is een tropische ziekte die de schistosoma-platwormen veroorzaken. Dat zijn parasieten die in de bloedvaten van de mens leven en zich daar voeden met rode bloedcellen. Ze leggen 300 eitjes per dag die deels via de urine of ontlasting het lichaam verlaten. De eitjes die in het water terechtkomen, ontluiken in larven (miracidia) die een zoetwaterslak als tussengastheer gebruiken. Deze slakken zijn een soort broedkamer waarin de larven zich ontwikkelen en zich ongeslachtelijk voortplanten. De honderdduizenden gekloonde larven (cercaria) verlaten de slak en zetten de jacht in op hun laatste gastheer: de mens. Als de mens zich in besmet water begeeft, dringen de cercaria door de huid het lichaam binnen. Ze scheiden enzymen af die de huid doordringbaar maken. Daarna migreren ze langs het bloedvatenstelsel via de longen en het hart naar de leverpoortader. Van daaruit vertrekken de volwassen wormen naar de bloedvaten rond de urineblaas (urinaire bilharzia) of rond de darmen (intestinale bilharzia), waar ze beginnen aan hun voortplanting. De cyclus begint dan opnieuw.

(http://eoswetenschap.eu/artikel/tropische-ziektes-veroveren-europa)

A. De auteur maakt een vergelijking.

B. De auteur onderbouwt een stelling met argumenten.

C. De auteur beschrijft een opeenvolging in tijd.

D. De auteur stelt een probleem vast en zoekt een oplossing.

GULS3

Zet onderstaande zinnen in de juiste volgorde.

[F] Hoe de dinosauriërs uitstierven, is nog steeds niet helemaal opgehelderd.

[B] Een populaire theorie is dat de meteorietinslag in Mexico een grote hoeveelheid zwavelzuurdeeltjes de lucht in katapulteerde.

[E] Die zouden het zonlicht tegengehouden hebben, zodat het overal ter wereld koud en donker werd.

[C] Die hypothese klopt niet.

[A] Op een ijskoude, donkere aardbol zouden veel meer soorten uitgestorven zijn, zoals de krokodillen.

[D] Bovendien betwijfelen wetenschappers of zwavelzuurdeeltjes wel zo lang in de lucht kunnen blijven na een meteorietinslag.

(http://eoswetenschap.eu/artikel/dinosauri-rs-gingen-rook-op)

GULS4

Zet onderstaande zinnen in de juiste volgorde.

[B] Eicellen worden in het lichaam blootgesteld aan schadelijke afbraakproducten, waaronder vrije radicalen.

[A] Dit zijn reactieve deeltjes die gemakkelijk schade aan kunnen richten in cellen.

[E] Hoe langer een eicel in het lichaam heeft gezeten, hoe meer beschadigingen er zijn opgetreden.

[C] Daardoor raken eicellen die op latere leeftijd vrijkomen minder makkelijk bevrucht.

[D] Ook slagen ze er bijvoorbeeld niet in om zich te nestelen in de baarmoeder waardoor de bevruchting uitloopt op een vroeg miskraam.

(http://www.kennislink.nl/publicaties/hoera-voor-de-oudere-moeder)

GULS5

Resultaten van deze studie toonden aan dat jongere mensen met een hoge socio-economische status (SES) significant meer sporten dan hun oudere metgezellen met een lage SES. Dit is in lijn met ander onderzoek rond kenmerken van sportparticipatie bij de algemene bevolking (Trost et al., 2002). Andere kenmerken zijn het verschil tussen geslacht en etniciteit, waarbij normaal gezien mannen en autochtonen meer sporten dan vrouwen en etnisch-culturele minderheden. Markant in dit onderzoek is dat we geen algemeen verschil zagen tussen mannen en vrouwen, of tussen autochtonen en etnisch-culturele minderheden. Het verschil van geslacht en etniciteit werd volledig verklaard door de lage sportparticipatie van vrouwen van etnisch-culturele afkomst. Deze resultaten zetten nog eens in de verf dat initiatieven, die inzetten op het mobiliseren van deze lage socio-economische groep van etnisch-cultureel diverse afkomst, echt nodig zijn.

(Marlier, M., & Willem, A. (2014). Sport als middel tot integratie van etnisch culturele minderheden en de lagere sociale klasse: resultaten van een studie bij Buurtsport Antwerpen. MOMENTEN (BRUSSEL), (12), 82–86)

Welke stelling kan je uit deze tekst afleiden?

A. Allochtone mannen sporten in het algemeen meer dan vrouwen.
B. In het onderzoek bleek een grote groep allochtone vrouwen weinig te sporten.
C. De studie toont aan dat jonge, laagopgeleide mensen significant meer sporten dan oudere, hoogopgeleiden.

GULS6

Productie van lekkere en gezonde voeding start reeds vanaf de grondstoffen, waarbij wordt gestreefd naar een excellente beginkwaliteit. Tijdens de verwerking wordt meer en meer gebruik gemaakt van mildere conserveringsmethoden (bijvoorbeeld: mildere hittebehandelingen) om deze beginkwaliteit zo goed mogelijk te behouden. In vergelijking met klassieke conserveringsmethoden, resulteert dit echter in vele gevallen in afgewerkte producten met een mindere stabiliteit tijdens bewaring, waarbij zowel microbiologische als chemische afbraakprocessen kunnen optreden. Deze trend naar mildere behandelingen heeft ertoe geleid dat er hogere eisen aan verpakkingen worden gesteld, daar uiteindelijk zij tijdens bewaring deze afbraakprocessen kunnen vertragen.

(*Ragaert, P. (2010). Trends in verpakkingen. FOOD SCIENCE AND LAW, 1(1), 17–20*)

Welke stelling kan je uit deze tekst afleiden?

A. Goede verpakkingen zijn tegenwoordig belangrijker dan de beginkwaliteit van een product.
B. Hoe milder de conserveringsmethoden zijn tijdens de verwerking van voedingsmiddelen, hoe beter de verpakkingen moeten zijn.
C. Milde conserveringsmethoden zoals een mildere hittebehandeling vertragen de afbraakprocessen in voedingsmiddelen.

GULS7

IJzer speelt een belangrijke rol in ons lichaam. Zo zit het in veel eiwitten en is het betrokken bij de opname van zuurstof in ons bloed. Maar als dit metaal in contact komt met UV-straling, wordt het erg instabiel. Dit instabiele ijzer zorgt ervoor dat er zuurstofradicalen ontstaan, die met alles in de cel gaan reageren. Vooral de mitochondriën, de energiecentrales van de cel, hebben hieronder te lijden. Mitochondriën bevatten veel ijzer en zuurstof, dus die gaan al snel kapot door de radicalen. Zonder energiecentrale is de rest van de cel verloren en al snel sterft deze. *(http://www.kennislink.nl/publicaties/zonnebrand-in-de-cel)*

Welke stelling kan je uit deze tekst afleiden?

A. Als het lichaam te veel ijzer opneemt, kunnen lichaamscellen afsterven.
B. Zuurstofradicalen ontstaan als er te veel zuurstof wordt opgenomen in het bloed.
C. UV-straling maakt het ijzer in ons lichaam instabiel.

GULS8

Als ouders hun kinderen steeds minder in dialect opvoeden, waarin voeden ze hun kinderen dan wel op? Uit het onderzoek van Soete (2012) blijkt dat vele ouders hun kinderen in standaardtaal trachten op te voeden, aangezien dat belangrijk zou zijn voor de carrièrekansen van het kind. Dat dat gerapporteerde gedrag ook overeenstemt met het werkelijke gedrag, lijkt echter weinig waarschijnlijk. Een goede standaardtaalbeheersing is voor veel West-Vlamingen immers nog steeds geen evidentie, aldus ook Vandekerckhove (2000). Bovendien staat standaardtaal ook heel ver af van het dialect, de informele omgangstaal waarin de meeste ouders zelf opgevoed zijn.

(*Ghyselen, A.-S. (2012). West-Vlaams is hot, of niet? OVER TAAL (KORTRIJK-HEULE), 51, 101–103*)

Welke stelling kan je uit deze tekst afleiden?

A. Uit het onderzoek van Soete (2012) blijkt dat veel West-Vlaamse ouders hun kinderen in standaardtaal opvoeden.
B. In het onderzoek van Soete (2012) zeiden veel West-Vlaamse ouders dat ze hun kinderen in standaardtaal opvoeden.
C. Dit onderzoek bewijst dat West-Vlaamse ouders vooral dialect spreken met hun kinderen.

GULS9

Welk woord past in beide onderstaande contexten?

A. omkeerbaar
B. irreversibel
C. tijdelijk

- Het aantal zenuwcellen voor een individu ligt bij de geboorte reeds vast voor de rest van het leven. Dit betekent dat beschadigde zenuwcellen zichzelf niet meer kunnen herstellen, en dat die beschadiging dus leidt tot een ……………………………………………… letsel.

- In de fysica spreken we van een …………………………………………………
  proces wanneer we bijvoorbeeld warm water bij koud water voegen. Na een tijdje is de temperatuur van het water overal dezelfde en kunnen we niet meer de ene helft koud en de andere helft warm maken.

GULS10

Welk woord past in beide onderstaande contexten?

A. opponent

B. oppositie

C. tegenligger

- Om in een debat de ander te overtuigen van je gelijk, moet je ervoor zorgen dat je ………………………………………………… je niet in de verdediging drukt of je ter verantwoording oproept.
- Tijdens het EK 2016 hebben de Rode Duivels hun ………………………………………………… Wales erg onderschat. Of dat al dan niet aan de bondscoach lag, laten we liever in het midden.

GULS11

Welk woord past in beide onderstaande contexten?

A. contrasteerde

B. contempleerde

C. contesteerde

- Verwoerd was niet zo tuk op de noemer ‘apartheid’, hij sprak namelijk liever over ‘goed nabuurschap’. Op die manier ……………………………………………… hij de betekenis van wat zich afspeelde in de Zuid-Afrikaanse samenleving.
- De feministe Yosefa Yoteyko ……………………………………………… de gedachte niet dat er meer mannelijke dan vrouwelijke genieën waren – dat gold in de toenmalige wetenschap als een vaststaand feit – maar wel dat daaruit algehele vrouwelijke intellectuele minderwaardigheid zou volgen.

GULS12

Vul het juiste signaalwoord in.

A. Toch
B. Bovendien
C. Zo

Een aantal plantensoorten beschikt over aanpassingen die langeafstandsverbreiding mogelijk maken (bijvoorbeeld: windverbreiders). ……………………………………………… zijn de afstanden waarover de meeste soorten verbreiden via zaad beperkt.

(Naar: D’hondt, B., Breyne, P., Van Landuyt, W., & Hoffmann, M. (2012). Draadklaver ontrafeld: de opmerkelijke rol van de mens als verbreider. NATUUR.FOCUS, 11(3), 112–118)

GULS13

Vul het juiste signaalwoord in.

A. daarentegen
B. weliswaar
C. immers

Grote mensenmassa's die zich voortbewegen over relatief kleine oppervlaktes vormen voor organisatoren een grote uitdaging. Organisatoren moeten …........................... te allen tijde de veiligheid kunnen garanderen.

(Versichele, M., Neutens, T., Huybrechts, R., Vlassenroot, S., & Gautama, S. (2012). Bluetooth: meer dan gadget voor mobiliteitsonderzoek. VERKEERSSPECIALIST (MECHELEN), (192), 26–29)

GULS14

Duid de juiste omschrijving aan voor het signaalwoord in vet.

A. Op voorwaarde dat

B. Hoewel

C. Zonder dat

Om deze laatste ondervindingen verder te ondersteunen, ging dit onderzoek na of de hypothese ook in de omgekeerde richting opgaat, namelijk een trager herstel van de oogzenuw wanneer er inflammatie gereduceerd wordt. Deze laatste experimenten van de thesis toonden echter geen effect aan. **Mits** er aan de experimentele werkwijze wordt gesleuteld, zou er wel een effect kunnen worden vastgesteld.

(Naar: An Beckers (2005). De invloed van acute inflammatie en inflammaging op het regeneratief potentieel van de zebravisretina, KULeuven)

GULS15

Kies het juiste woord.

Een ……………………………………… is een vermogen dat kennis, inzicht, attitudes en vaardigheden bundelt om in concrete taaksituaties doelen te bereiken.

A. **competentie**
B. competitie
C. competent

GULS16

Kies het juiste woord.

Ik heb er …………………………………… op aangedrongen bij de directeur om de zaak te herbekijken, maar ze weigert op haar besluit terug te komen.

A. **herhaaldelijk**
B. herhaald
C. herhalend

GULS17

Kies het juiste woord.

De nieuwe voorzitter moet een …………………………………..…………………. reputatie hebben. Als je zo een belangrijke positie bekleedt, kun je geen schandalen gebruiken.

A. omstreden
B. **onomstreden**
C. onstreden

GULS18

Kies het juiste woord.

Het hof erkent dat dit soort regels nodig kunnen zijn in de praktijk, en dat ze niet altijd in strijd zijn met het verdrag, maar vinden het feit dat geen afwijking mogelijk is 'buitensporig star en …………………………………….………. tegenover vrouwen'.

A. **discriminatoir**
B. discriminatair
C. discriminair

GULS19

Kies het juiste woord.

Het Hof besluit tot een schending van artikel 14 samen met artikel 8, omdat de bepaling van de naam van wettelijke kinderen 'enkel en alleen gemaakt werd op basis van een discriminatie gebaseerd op het geslacht van de ouders'. Daarnaast erkent het Hof dat dit soort regels gegrond zijn in een ………………………………………………… opvatting van de familie en de macht van de echtgenoot.

A. patriarchische
B. **patriarchale**
C. patriottische

GULS20

Kies het juiste woord.

Zijn vader was een nogal ………………………………………………………. figuur. Hij had lak aan alle regels en dat bracht hem vaak in conflict met zijn omgeving.

A. inconventioneel

B. onconventioneel

C. aconventioneel

GULS21 tot en met GULS25

Plaats de juiste woordgroep in onderstaande zinnen.

A. staan afwijzend tegenover

B. zijn gebaat bij

C. gaan gepaard met

D. staan haaks op

E. sluiten naadloos aan bij

De ideeën van de communistische partij ……………………………………… (*staan haaks op*) de uitgangspunten van een neoliberaal systeem.

De resultaten van de experimenten uit het onderzoek van professor Vandenberg ……………………………………… (*sluiten naadloos aan bij*) de bevindingen van ons onderzoeksteam. Dat wil zeggen dat beide onderzoeksinitiatieven de hypothese van professor Vandenberg bevestigen.

Drugs ……………………………………… (*gaan gepaard met*) negatieve bijwerkingen.

Zowel automobilisten als fietsers ……………………………………… (*zijn gebaat bij*) goed werkende fietslichten.

Hoewel het wetsvoorstel positief werd onthaald in het parlement, wil de minister het niet aanvaarden. De reden: alle partijleiders ……………………………………… (*staan afwijzend tegenover*) het voorstel.