

English follows Dutch

Kennisgeving dataverzameling voor REACH: Research on Educational Success, Academic Language, and Curricula in Higher Education

Met dit bericht willen we alle UvA-studenten en recente alumni informeren over het REACH-project en over welke soorten studentgegevens wij in dat project zullen gebruiken. Het onderzoeksproject REACH (Research on Educational Success, Academic Language and Curricula in Higher Education) maakt namelijk gebruik van gearchiveerde studentgegevens in bestaande UvA-databases. Het project wordt uitgevoerd door promovendi Agnieszka Kluska en Afon Khari onder begeleiding van prof. dr. Sible Andringa aan de Universiteit van Amsterdam, Amsterdam Center for Language and Communication.

Doel van het onderzoeksproject

Studenten in het hoger onderwijs variëren in hun taalachtergrond en taalvaardigheid, en ook in de mate waarin ze succesvol hun studie doorlopen. Eerder onderzoek laat zien dat de taalvaardigheid van studenten een belangrijke rol kan spelen in hun studieprestaties, bijvoorbeeld in de cijfers die zij halen en hoe lang zij over hun studie doen. Toch weten we heel weinig over hoe dit werkt in het Nederlandse hoger onderwijs. Dit project onderzoekt daarom hoe academische taalvaardigheid zich in de loop van de tijd ontwikkelt, of curricula voldoende ondersteuning bieden, en welke factoren van belang zijn voor studiesucces. Om dit te onderzoeken, kijken we eerst naar patronen in studentgegevens uit eerdere jaren. Dit betekent dat we gegevens opvragen uit UvA-databases, met name informatie over studentachtergronden en studiesucces, evenals schrijfvaardigheid. We zullen kwantitatieve analysemethoden gebruiken om verbanden tussen variabelen onderzoeken. We willen bijvoorbeeld weten of studenten die meer schrijfondersteuning tijdens hun studie ontvangen ook hogere cijfers halen, welk type taalondersteuning het meest bijdraagt aan studiesucces, en of het uitmaakt of studenten in hun moedertaal studeren. Met dit project willen we het taalbeleid helpen vormen en opleidingen wetenschappelijk onderbouwde handvatten bieden om studenten sneller, met meer plezier en succes te laten studeren.

Welke gegevens worden verzameld?

Deze projecten richten zich zowel op studentkenmerken als op het geschreven werk van bachelorstudenten. We gebruiken gearchiveerde gegevens uit de databases van de Universiteit van Amsterdam van (maximaal) de academische jaren 2018/19 – 2028/29. Hoe groot de uiteindelijke dataset wordt, hangt af van welke gegevens echt beschikbaar zijn; de dataset zal maximaal ongeveer 4000 studenten per cohort (afstudeerlichting per jaar) omvatten.

We zullen 9 verschillende maar samenhangende soorten gegevens uit de UvA-databases opvragen.

1. Bachelorscripties met cijfers en beoordelingsformulieren (rubrics).
2. Een selectie van eerdere individuele schrijfpodrachten, zoals essays, onderzoeksvorstellen en labverslagen, met cijfers en beoordelingsformulieren (rubrics).

Deze documenten worden voorbereid en gepseudonimiseerd voordat ze geanalyseerd worden. Dit betekent dat alle studentgegevens, zoals namen, studentnummers en e-mailadressen, maar ook titelpagina's en dankwoorden, worden verwijderd. Door naar schrijfproducten van studenten te kijken, kunnen we patronen in taalgebruik identificeren en zien of deze verschillen tussen disciplines of tussen moedertaalsprekers en niet-

moedertaalsprekers. Met die kennis kunnen we vervolgens bepalen welke groepen studenten met welke vormen van taalondersteuning het beste geholpen kunnen worden.

3. Taalachtergrondinformatie.

Je kunt je misschien voorstellen dat de taal of talen waarmee je bent opgegroeid invloed zou kunnen hebben op je taalvaardigheid. De UvA slaat dit soort informatie echter niet op, en daarom zullen wij deze informatie schatten. Eerst zullen we nagaan of studenten bij inschrijving in hun opleiding een bewijs van taalvaardigheid moesten aanleveren, zoals TOEFL- of IELTS-resultaten voor een Engelstalige opleiding. De testresultaten zelf verwerken we niet, alleen of dit bewijs is geaccepteerd. Ook zullen we de nationaliteit en het geboorteland van studenten tijdelijk gebruiken, maar **uitsluitend** in combinatie met andere beschikbare informatie om een inschatting te kunnen maken van de taalachtergrond van een student. Daarna worden deze gegevens uit onze dataset verwijderd. We volgen hierbij het principe van dataminimalisatie: zo weinig mogelijk gegevens gebruiken, gedurende een zo kort mogelijke periode.

4. Demografische informatie, specifiek leeftijd en gender.
5. Gegevens over studieprestaties en studievoortgang, zoals cijfers, behaalde ECTS, studieduur, diplomadatum, studie-uitval, studievertraging en andere indicatoren van studiesucces.
6. Opleidings- en curriculumgerelateerde informatie, zoals opleiding, track, opleidingsvoertaal, cohort, jaar van inschrijving en relevante opleidingskenmerken.
7. Onderwijsachtergrond/-geschiedenis, zoals vooropleiding, type diploma, eerdere instelling, of land van eerdere kwalificatie, voor zover nodig en beschikbaar.

Uit eerder onderzoek en ervaring weten we dat studentpopulaties sterk kunnen verschillen tussen opleidingen en onderwijssituaties, en dat dit van belang is voor het begrijpen van verschillen in academische trajecten, deelnamepatronen en beleidsimplicaties. Daarom willen we studieresultaten koppelen aan leeftijd en gender, en aan informatie over de opleiding die tot deze resultaten heeft geleid. Een identificatienummer helpt daarbij, en daarom verwerken we ook gegevens daarvoor:

8. De UvA zelf gebruikt o.a. studentnummers om studentgegevens te koppelen en te pseudonimiseren. Wij zullen echter project-specifieke identificatienummers gebruiken in plaats van direct herleidbare studentnummers.

We verkrijgen gegevens uit de administratieve systemen van de Universiteit van Amsterdam via daartoe gemachtigde UvA-dataservices. We verwerken en bewaren de gegevens uitsluitend binnen veilige, door de UvA goedgekeurde onderzoeksomgevingen.

9. Tot slot verzamelen we, naast gearchiveerde studentgegevens, informatie over de curricula van bacheloropleidingen aan de Universiteit van Amsterdam. We verzamelen deze informatie van openbaar toegankelijke universitaire websites, zoals de studiegids en de Onderwijs- en Examenregelingen (OER'en). We richten ons op taalgerelateerde factoren, zoals of de opleiding diagnostische taaltoetsen afneemt, hoe opleidingen taal beoordelen, en of er taalgerelateerde eindkwalificaties zijn. Dit zijn geen persoonsgegevens, maar gegevens over de opleidingen waarin studenten zich bevinden.

Wat als je dit niet wilt (opt-out)?

Als je liever niet hebt dat je gegevens in dit onderzoek worden opgenomen, is dat geen probleem. Niet alle studentgegevens zullen worden meegenomen, maar als je zeker wilt weten dat je gegevens niet worden gebruikt in dit onderzoek, kun je dat eenvoudig aan ons laten weten. Jouw gegevens zullen dan niet worden ingezien of gebruikt. Als je uitgesloten wilt

worden van het onderzoek, neem dan contact op met een van de promovendi, Agnieszka Kluska of Afon Khari (e-mail: a.l.kluska@uva.nl, m.khari3@uva.nl, Spuistraat 134, 1012 VB Amsterdam). Vermeld in de onderwerpregel van je bericht "Exclusion from REACH" en zet je studentnummer in het bericht. Op die manier kunnen wij de bijbehorende informatie in de dataset terugvinden en verwijderen. Je ontvangt vervolgens een bevestiging dat dit is gebeurd en daarna verwijderen wij ook de correspondentie.

Ongemak, risico's & verzekering

Er zijn geen vormen van ongemak of risico's verbonden aan deelname. De gegevens worden niet gedeeld met derden. Zoals voor al het onderzoek aan de Universiteit van Amsterdam, geldt er een standaard aansprakelijkheidsverzekering.

Rechten van betrokkenen volgens de AVG

Als je meer wilt weten over je rechten als betrokkene onder de Europese privacywetgeving, de AVG (GDPR), kijk dan op deze website: <https://extranet.uva.nl/en/content/a-z/privacy/what-are-my-rights/what-are-my-rights.html>.

Vertrouwelijke behandeling van persoonsgegevens

Uiteraard willen we je gegevens goed beveiligen. De tijdens dit onderzoek verzamelde informatie wordt uitsluitend gebruikt voor dit onderzoeksproject. Persoonsgegevens zullen niet in publicaties worden gebruikt. De gegevens worden gepseudonimiseerd en de encryptiesleutel wordt op een veilige en aparte locatie bewaard, die alleen toegankelijk is voor leden van het onderzoeksteam.

De persoonsgegevens worden alleen opgeslagen zolang dat nodig is voor het onderzoek en worden zo snel mogelijk verwijderd. Overige gegevens (niet-persoonsgegevens) worden voor onbepaalde tijd bewaard, conform de vereisten voor promotietrajecten.

Nadere informatie

Als u meer wilt weten over het onderzoeksproject, kunt u contact opnemen met de promovendi Agnieszka Kluska, Afon Khari en/of de begeleider prof. dr. Sible Andringa (e-mail: a.l.kluska@uva.nl, m.khari3@uva.nl, s.j.andringa@uva.nl; Spuistraat 134, 1012 VB Amsterdam).

Als u klachten heeft over dit onderzoeksproject, kunt u contact opnemen met de secretaris van de Ethiekcommissie van de Faculteit der Geesteswetenschappen van de Universiteit van Amsterdam, commissie-ethiek-fgw@uva.nl; Binnengasthuisstraat 9, 1012 ZA Amsterdam, Nederland.

English

Data collection notice for REACH: Research on Educational Success, Academic Language, and Curricula in Higher Education

This message is to let you know about the REACH project and what kinds of student data we will be using in that project. REACH (Research on Educational Success, Academic Language and Curricula in Higher Education) is a research project that makes use of archived student data in existing UvA databases. The project is being carried out by PhD students Agnieszka Kluska and Afon Khari under supervision of prof. dr. Sible Andringa at the University of Amsterdam, Amsterdam Centre for Language and Communication.

Purpose of the research project

Higher education is linguistically diverse but academic success is not equally distributed. Previous research shows that language abilities of students can be an important factor in their academic performance, e.g., what grades they obtain and how long they take to graduate, yet we know very little about how this works in Dutch higher education. This project therefore looks into how academic language skills develop over time, whether program curricula offer sufficient support, and which factors matter for academic success. To do so, we will first look for patterns in student data from previous years. This means we will retrieve data from university databases, especially information relating to student backgrounds and study success, as well as writing skills. We will use quantitative methods of analysis to establish relationships between variables, e.g., do students with more writing support in their programs achieve higher grades on their assignments? Does it differ for native and non-native English/Dutch speakers? What type of language support is the most effective and helps students graduate on time? Our goal is to contribute to evidence-informed language policy and curricula at the University of Amsterdam in order to help students study more happily, quickly and successfully.

What data is collected?

These projects look at both student characteristics as well as students' written work in bachelor programs. We will use archived data from the databases at the University of Amsterdam spanning (at most) the academic years 2018/19 – 2028/29. The size of the dataset depends on data availability, and would amount to data from around 4000 students at most per cohort, i.e., the graduating class each year.

We plan to retrieve 9 different but related types of data from the UvA databases.

1. Bachelor theses with grades and assessment forms (rubrics).
2. A selection of earlier individual long-written assignments, such as essays, research proposals and lab reports, with grades and assessment forms (rubrics).

These documents will be pre-processed and pseudonymized prior to any analysis. This means that all mentions of a student's name, student ID or e-mail address, as well as the thesis title pages and acknowledgments will be deleted. By looking at student writing, we can identify patterns of language use, and see whether they differ across disciplines, or between native and non-native speakers. Knowing this, we can then see which groups of students might be helped by what types of language support.

3. Language-background information.

You can imagine that a student's language skills might depend on what language(s) they use or grew up with. The UvA does not store this kind of information, so we will estimate this information instead. First, we will see whether students were asked to prove their language proficiency when they enrolled in the program they were admitted to, like TOEFL or IELTS results for an English-taught program. We will not process the test results themselves, only whether that proof was accepted. We will also temporarily use students' nationality and country of birth, but **only** in combination with other available information to estimate a student's linguistic background. Once we have done so, the data that we used for this inference will be deleted from our dataset. We follow the principle of data minimization here: use as little as possible, for as short a time as possible.

4. Demographic information, specifically age and gender.
5. Academic achievement and study progress data, such as grades, ECTS completion, study duration, diploma completion, retention, delay, and other study-success indicators.
6. Program and curriculum-related information, such as program, track, language of instruction, cohort, year of enrolment, and relevant program characteristics.
7. Educational background/history, such as prior education, diploma type, previous institution, or country of prior qualification where available and relevant.

From prior research and experience we know that student populations can differ substantially across programs and educational contexts, and this matters for understanding differences in academic trajectories, program participation patterns, and policy implications. This is why we want to connect student results with their age and gender, and information about the program that led them to those results. A student identifier helps with that, which is why we also process 8:

8. The UvA uses student numbers and/or other administrative identifiers to link and pseudonymize data. We will use project-specific pseudonymous identifiers rather than directly identifying student numbers.

We retrieve identifier data from the University of Amsterdam administrative systems via authorized UvA data services. We will process and store the data only within secure UvA-approved research environments.

9. Finally, besides archival student data, we will collect information about the curricula of bachelor programs at the University of Amsterdam. We will gather it from publicly available university websites, such as the Course Catalogue and the Teaching and Examination Regulations. We will focus on language-related factors, such as the presence of diagnostic language tests, methods of language assessment, and language-related exit qualifications. This is not personal data, but data about the program students are in.

What if you want to opt out?

If you would rather not have your data included in this study, that's ok. Not every student's data will be included, but if you would like to be sure that yours is not, let us know. You can easily opt out from the research project and your data will not be accessed nor used. If you wish to be excluded from the study, please contact one of the PhD students Agnieszka Kluska or Afon Khari (email: a.l.kluska@uva.nl, m.khari3@uva.nl, Spuistraat 134, 1012VB Amsterdam). Please title your e-mail "Exclusion from REACH" and include your student ID number in the message. This way we can find the information associated with you in the dataset, and delete it. You will then receive confirmation that this has happened and we will delete the correspondence as well.

Discomfort, Risks & Insurance

There is no discomfort or risks involved. The data will not be shared with any unauthorised parties. For all research at the University of Amsterdam, a standard liability insurance applies.

Data subject rights according to the GDPR

If you want to know more about your rights as data subjects under the EU privacy law, the GDPR, please see this website: <https://extranet.uva.nl/en/content/a-z/privacy/what-are-my-rights/what-are-my-rights.html>.

Confidential treatment of personal data

Of course we want to keep your data safe. The information gathered over the course of this research will be used for the purpose of this research project. Personal details will not be used in publications. The data will be pseudonymised, and the encryption key will be stored in a secure and separate location, and accessible only to members of the research project team.

The personal data will only be stored as long as is necessary for the research and will be deleted as soon as possible. Remaining data (non-personal data) will be stored indefinitely, as per requirements for PhD projects.

Further information

If you'd like to know more about the research project, please contact the PhD students Agnieszka Kluska, Afon Khari and/or the supervisor prof. dr. Sible Andringa (email: a.l.kluska@uva.nl, m.khari3@uva.nl, s.j.andringa@uva.nl; Spuistraat 134, 1012VB Amsterdam).

If you have any complaints regarding this research project, you can contact the secretary of the Ethics Committee of the Faculty of Humanities of the University of Amsterdam, commissie-ethiek-fgw@uva.nl; Binnengasthuisstraat 9, 1012 ZA Amsterdam, The Netherlands.