

## University of Groningen

### Implementing assessment innovations in higher education

Boevé, Anna Jannetje

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*  
2018

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Boevé, A. J. (2018). *Implementing assessment innovations in higher education*. [Thesis fully internal (DIV), University of Groningen]. Rijksuniversiteit Groningen.

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

# R

## References



- Abeysekera, L., & Dawson, P. (2015). Motivation and cognitive load in the flipped classroom: definition, rationale and a call for research. *Higher Education Research and Development, 34*(1), 1-14. doi:10.1080/07294360.2014.934336
- Alexander, F. K. (2000). The changing face of accountability: Monitoring and assessing institutional performance in higher education. *The Journal of Higher Education, 71*, 411-431. doi:10.2307/2649146
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington: American Educational Research Association.
- Anakwe B. (2008). Comparison of student performance in paper-based versus computer-based testing. *Journal of Education for Business, 84*(1), 13-17. doi: 10.3200/JOEB.84.1.13-17
- Angus, S. D., & Watson, J. (2009). Does regular online testing enhance student learning in the numerical sciences? Robust evidence from a large data set. *British Journal of Educational Technology, 40*(2), 255-272. doi:10.1111/j.1467-8535.2008.00916.x
- Aojula, H., Barber, J., Cullen, R., & Andrews, J. (2006). Computer-based, online summative assessment in undergraduate pharmacy teaching: The Manchester experience. *Pharmacy Education, 6*(4), 229-236. doi:10.1080/15602210600886209
- Apostolou, B., Blue, M. A., & Daigle, R. J. (2009). Student perceptions about computerized testing in introductory managerial accounting. *Journal of Accounting Education, 27*(2), 59-70. doi:10.1016/j.jaccedu.2010.02.003
- Ashwin, P., & McVitty, D. (2015). The meanings of student engagement: implications for policies and practices. In *The European Higher Education Area* (pp. 343-359). Springer International Publishing. doi:10.1007/978-3-319-20877-0\_23
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01
- Battaaz, M. (2015). equateIRT: An R package for IRT test equating. *Journal of Statistical Software, 68*(7), 1-22. doi:10.18637/jss.v068.i07
- Bayazit A., & Askar, P. (2012). Performance and duration differences between online and paper-pencil tests. *Asia Pacific Educational Review, 13*(2), 219-226. doi: 10.1007/s12564-011-9190-9
- Beatty, A. S., Walmsley, P. T., Sackett, P. R., & Kuncel, N. R. (2015). The reliability of college grades. *Educational Measurement: Issues and Practice, 34*(4), 31-40. doi:10.1111/emip.12096
- Biggs, J., Kember, D., & Leung, D. Y. (2001). The revised two-factor study process questionnaire: R-SPQ-2F. *British Journal of Educational Psychology, 71*(1), 133-149. doi:10.1348/000709901158433
- Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-74. doi:10.1080/0969595980050102
- Black, P., & William, D. (2003). 'In praise of educational research': Formative assessment. *British Educational Research Journal, 29*(5), 623-637. doi:10.1080/0141192032000133721

- Boevé, A. J., Meijer, R. R., Albers, C. J., Beetsma, Y., & Bosker, R. J. (2015). Introducing computer-based testing in high-stakes exams in higher education: Results of a field experiment. *PLoS one*, *10*(12), doi:10.1371/journal.pone.0143616
- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., ... & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, *86*(4), 803-848. doi:10.3102/0034654316672069
- Bussemaker, J. (2014). Brief aan de voorzitter van de Tweede kamer der Staten Generaal: Voortgang Hoofdlijnenakkoorden en Prestatieafspraken Hoger Onderwijs en Onderzoek. The Hague, Netherlands: Ministerie van Onderwijs, Cultuur en Wetenschap
- Burnard, P., Gill, P., Stewart, K., Treasure, E., & Chadwick, B. (2008). Analysing and presenting qualitative data. *British Dental Journal*, *204*(8), 429-432. doi:10.1038/sj.bdj.2008.292
- Cagiltay, N., & Ozalp-Yaman, S. (2013). How can we get benefits of computer-based testing in engineering education?. *Computer Applications in Engineering Education* *21*(2), 287-293. doi:10.1002/cae.20470
- Cai, L., Thissen, D., & du Toit, S. H. C. (2011). IRTPRO for windows [computer software]/ Scientific Software International, Lincolnwood, IL
- Cantillon, P., Irish, B., & Sales, D. (2004). Using computers for assessment in medicine. *BMJ*, *329*, 606-609. doi:10.1136/bmj.329.7466.606
- Carey, T. A., & Stiles, W. B. (2015). Some problems with randomized controlled trials and some viable alternatives. *Clinical Psychology & Psychotherapy*, *23*(1), 87-95. doi:10.1002/cpp.1942
- Carpenter, S. K., Rahman, S., Lund, T. J., Armstrong, P. I., Lamm, M. H., Reason, R. D., & Coffman, C. R. (2017). Students' Use of Optional Online Reviews and Its Relationship to Summative Assessment Outcomes in Introductory Biology. *CBE-Life Sciences Education*, *16*(2), ar23. doi:10.1187/cbe.16-06-0205
- Carrillo-de-la-Peña, M. T., Bailles, E., Caseras, X., Martínez, À., Ortet, G., & Pérez, J. (2009). Formative assessment and academic achievement in pre-graduate students of health sciences. *Advances in Health Sciences Education*, *14*(1), 61-67. doi:10.1007/s10459-007-9086-y
- Cizek, C. K. (2009). Reliability and validity of information about student achievement: Comparing large scale and classroom testing effects. *Theory into Practice*, *48*(1), 63-71. doi:10.180/004058408408025.77627
- Clariana, R., & Wallace, P. (2002). Paper-based versus computer-based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology*, *33*(5), 593-602. doi:10.1111/1467-8535.00294
- Credé, M., & Phillips, L. A. (2011). A meta-analytic review of the Motivated Strategies for Learning Questionnaire. *Learning and Individual Differences*, *21*(4), 337-346. doi:10.1016/j.lindif.2011.03.002
- Credé, M., Roch, S. G., & Kieszczynka, U. (2010). Class attendance in college: A meta-analytic review of the relationship of class attendance with grades and student characteristics. *Review of Educational Research*, *80*(2), 272-295. doi:10.3102/0034654310362998

- Cronbach, L. J. (1977). *Essentials of Psychological Testing*. New York, Harper & Brothers.
- Csapo, B., Ainley, J., Bennett, R. E., Latour, T., & Law N. (2012). Technological issues for computer-based assessment. Griffin P., McGaw B., Care E. (eds). *Assessment and teaching of 21st century skills*. Dordrecht, Netherlands: Springer, pp. 143
- Davies, R. S., Dean, D. L., & Ball, N. (2013). Flipping the classroom and instructional technology integration in a college-level information systems spreadsheet course. *Educational Technology Research and Development*, 61(4), 563-580. doi:10.1007/s11423-013-9305-6
- De Boer, H. F., Jongbloed, B. W. A., Benneworth, P. S., Cremonini, L., Kolster, R., Kottmann, A., ... & Vossensteyn, J. J. (2015). Performance-based funding and performance agreements in fourteen higher education systems (Report for the Ministry of Education, Culture, and Science). Enschede, Netherlands: Center for Higher Education Policy Studies.
- De Kleijn, R. A., Bouwmeester, R. A., Ritzen, M. M., Ramaekers, S. P., & Van Rijen, H. V. (2013). Students' motives for using online formative assessments when preparing for summative assessments. *Medical Lecturer*, 35(12), e1644-1650. doi:0.3109/0142159X.2013.826794
- Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *Educational Psychologist*, 26(3-4), 325-246. doi:10.1080/00461520.1991.9653137
- Dermo, J. (2009). e-Assessment and the student learning experience: A survey of student perceptions of e-assessment. *British Journal of Educational Technology*, 40(2), 203-214. doi:10.1111/j.1467-8535.2008.00915.x
- Deutsch, T., Herrmann, K., Frese, T., & Sandholzer, H. (2012). Implementing computer-based assessment—a web-based mock examination changes attitudes. *Computers & Education*, 58(4), 1068-1075. doi:10.1016/j.compedu.2011.11.013
- Dobson, J. L. (2008). The use of formative online quizzes to enhance class preparation and scores on summative exams. *Advances in Physiology Education*, 32(4), 297-302. doi:10.1152/advan.90162.2008
- Dollinger, S. J., Matyja, A. M., & Huber, J. L. (2008). Which factors best account for academic success: Those which college students can control or those they cannot? *Journal of Research in Personality*, 42(4), 872-885. doi:10.1016/j.jrp.2007.11.007
- Dove, A. (2013). Students' perceptions of learning in a flipped statistics class. In R. McBride & M. Searson (Eds.), *Proceedings of Society for Information Technology & Lecturer Education International Conference 2013* (pp. 393-398). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE).
- Draper, N. R., & Smith, H. (2014). *Applied regression analysis*. New York: John Wiley & Sons.
- El Shallaly, G. E., & Mekki, A. M. (2012). Use of computer-based clinical examination to assess medical students in surgery. *Educational Health*, 25(3), 148-152. doi:10.4103/1357-6283.109789
- Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107-115. doi:10.1111/j.1365-2648.2007.04569.x
- Embretson, S. E. & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah NJ: Erlbaum.

- Escudier, M. P., Newton, T. J., Cox, M. J., Reynolds, P. A., & Odell, E. W. (2011). University students' attainment and perceptions of computer delivered assessment; a comparison between computer-based and traditional tests in a 'high-stakes' examination. *Journal of Computer Assisted Learning*, 27(5), 440-447. doi:10.1111/j.1365-2729.2011.00409.x
- Fox, R. A., McManus, I. C., & Winder, B. C. (2001). The shortened Study Process Questionnaire: An investigation of its structure and longitudinal stability using confirmatory factor analysis. *British Journal of Educational Psychology*, 71(4), 511-530. doi:10.1348/000709901158659
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410-8415. doi:10.1073/pnas.1319030111
- Frein S. T. (2011). Comparing in-class and out-of-class computer-based tests to traditional paper-and-pencil tests in introductory psychology courses. *Teaching of Psychology*, 38(4), 282-287. doi: 10.1177/0098628311421331
- Fuentes-Pardo, J. M., García, A. I., Ramírez-Gómez, Á., & Ayuga, F. (2014). Computer-based tools for the assessment of learning processes in higher education: A comparative analysis. In 8<sup>th</sup> International Technology, Education and Development Conference Proceedings (pp. 976-984), Valencia, Spain
- Gibbs, G. (1999). Using assessment strategically to change the way students learn. In S. Brown & A. Glasner (Eds), *Assessment matters in higher education* (pp. 41–53) Buckingham: S.R.H.E. and Open University Press
- Grün, B., & Zeileis, A. (2009). Automatic Generation of Exams in R. *Journal of Statistical Software*, 29(10), 1-14, doi:10.18637/jss.v029.i10
- Haberman, S. J. (2008). When can subscores have value? *Journal of Educational and Behavioural Statistics*, 33(2), 204-229. doi: 10.3102/1076998607302636.
- Handelsman, M. M., Briggs, W. L., Sullivan, N., & Towler, A. (2005). A measure of college student course engagement. *The Journal of Educational Research*, 98(3), 184-192. doi:10.3200/JOER.98.3.184-192
- Harks, B., Klieme, E., Hartig, J., & Leiss, D. (2014). Separating Cognitive and Content Domains in Mathematical Competence. *Educational Assessment*, 19(4), 243-266. doi:10.1080/10627197.2014.964114
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. doi:10.3102/003465430298487
- Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research*, 66(2), 99-36. doi:10.3102/00346543066002099
- Hift, R. J. (2014). Should essays and other open ended type of questions retain a place in written summative assessment in clinical medicine? *BMC Medical Education*, 14(1), 249. doi:10.1186/s12909-014-0249-2.
- Hochlehnert, A., Brass, K., Moeltner, A., & Juenger, J. (2011). Does medical students' preference of test format (computer-based vs. paper-based) have an influence on performance?. *BMC Medical Education*, 11(1), 89. doi:10.1186/1472-6920-11-89

- Hollingshead, L., & Childs, R. A. (2011). Reporting the percentage of students above a cut-score: the effect of group size. *Educational Measurement: Issues and Practice*, 30(1), 36-43. doi:10.1111/j.1745-3992.2010.00198.x
- Hornsby, D. J., & Osman, R. (2014). Massification in higher education: Large classes and student learning. *Higher Education*, 67(6), 711-719. doi:10.1007/s10734-014-9733-1
- Huff, K. L., & Sireci, S. (2001). Validity issues in computer-based testing. *Educational Measurement: Issues and Practice*, 20(3), 16-25. doi: 10.1111/j.1745-3992.2001.tb00066.x
- Jensen, J. L., Kummer, T. A., & Godoy, P. D. D. M. (2015). Improvements from a flipped classroom may simply be the fruits of active learning. *CBE-Life Sciences Education*, 14(1), ar5. doi:10.1187/10.1187/cbe.14-08-0129
- Jeong, H. (2014). A comparative study of scores on computer-based tests and paper-based tests. *Behaviour & Information Technology*, 33(4), 410-422. doi:10.1080/0144929X.2012.710647
- Kahu, E. R. (2013). Framing student engagement in higher education. *Studies in Higher Education*, 38(5), 758-773. doi:10.1080/03075079.2011.598505
- Kalogeropoulos, N., Tzigounakis, I., Pavlatou, E.A., & Boudouvis, A.G. (2013). Computer-based assessment of student performance in programming courses. *Computer Applications in Engineering Education*, 21(4), 671-683. doi:10.1002/cae.20512
- Karpicke, J. D. (2009). Metacognitive control and strategy selection: deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, 138(4), 469. doi:10.1037/a0017341
- Kerdijk, W., Cohen-Schotanus, J., Mulder, B., Muntinghe, F. L., & Tio, R. A. (2015). Cumulative versus end-of-course assessment: effects on self-study time and test performance. *Medical Education*, 49(7), 709-716. doi:10.1111/medu.12756
- Kerdijk, W., Tio, R. A., Mulder, B. F., & Cohen-Schotanus, J. (2013). Cumulative assessment: strategic choices to influence students' study effort. *BMC Medical Education*, 13(1), 172. doi:10.1186/1472-6920-13-172
- Ketterlin-Geller, L. R., & Yovanoff, P. (2009). Diagnostic assessments in mathematics to support instructional decision making. *Practical Assessment, Research & Evaluation*, 14(16), 1-11. [no doi available]
- Kibble, J. (2007). Use of unsupervised online quizzes as formative assessment in a medical physiology course: effects of incentives on student participation and performance. *Advances in Physiology Education*, 31(3), 253-260. doi:10.1152/advan.00027.2007
- Kim, Y. H., & Goetz, E. T. (1993). Strategic processing of test questions: The test marking responses of college students. *Learning and Individual Differences*, 5(3), 211-218. doi: 10.1016/1041-6080(93)90003-B
- Kolen, M. J., & Brennan, R. L. (2014). Test equating, scaling, and linking: Methods and practices. New York: Springer-Verlag.
- Kostal, J. W., Kuncel, N. R., & Sackett, P. R. (2016). Grade inflation marches on: Grade increases from the 1990s to 2000s. *Educational Measurement: Issues and Practice*, 35(1) 11-20. doi:10.1111/emip.12077
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, 41(4), 212-218. doi:10.1207/s15430421tip4104\_2



- Kuh, G. D., Cruce, T. M., Shoup, R., Kinzie, J., & Gonyea, R. M. (2008). Unmasking the effects of student engagement on first-year college grades and persistence. *The Journal of Higher Education*, 79(5), 540-563. doi:10.1353/jhe.0.0019
- Lakens, D. (2017). Equivalence Tests: A practical primer for t tests, correlations and meta-analyses. *Social Psychological and Personality Science*, 8(4), 355-362. doi:10.1177/1948550617697177
- Lee G., & Weekaron P. (2001). The role of computer-aided assessment in health professional education: A comparison of student performance in computer-based and paper-and-pen multiple-choice tests. *Psychological Bulletin*, 23(2). doi:10.1080/01421590020031066
- Lievens, F. (2013). Adjusting medical school admission: Assessing interpersonal skills using situational judgment tests. *Medical Education*, 47(2), 182-189. doi:10.1111/medu.12089
- Lim, E. C., Ong, B. K., Wilder-Smith, E. P., & Seet, R. C. (2006). Computer-based versus pen-and-paper testing: Students' perception. *Annals of the Academy of Medicine Singapore*, 35(9), 599-603. [no doi available]
- Liu, O. L. (2011). Outcomes assessment in higher education: Challenges and future research in the context of voluntary system of accountability. *Educational Measurement: Issues and Practice*, 30(3), 2-9. doi:10.1111/j.1745-3992.2011.00206.x
- Lugtig, P. (2014). Panel attrition separating stayers, fast attriters, gradual attriters, and lurkers. *Sociological Methods & Research*, 43(4). 669-723. doi:10.1177/0049124113520305
- Lugtig, P., Glasner, T., & Boevé, A. J. (2015). Reducing underreports of behaviours in retrospective surveys: The effects of three different strategies. *International Journal of Public Opinion Research*. 28(4). 583-595. doi:10.1093/ijpor/edv032
- Luyten, H. (1994). Stability of school effects in Dutch secondary education: The impact of variance across subjects and years. *International Journal of Educational Research*, 21(2), 197-216. doi:10.1016/0883-0355(94)90032-9
- Marden, N. Y., Ulman, L. G., Wilson, F. S., & Velan, G. M. (2013). Online feedback assessments in physiology: effects on students learning experiences and outcomes. *Advances in Physiology Education*, 37(2), 192-200. doi:10.1152/advan.00092.2012
- Mason, G. S., Shuman, T. R., & Cook, K. E. (2013). Comparing the effectiveness of an inverted classroom to a regular classroom in an upper-division engineering course. *IEEE Transactions on Education*, 56(4), 430-435. doi:10.1109/TE.2013.2249066
- Maydeu-Olivares A., Kramp U., García-Forero C., Gallardo-Pujol D., & Coffman D. (2009). The effect of varying the number of response alternatives in rating scales: Experimental evidence from intra-individual effects. *Behaviour Research Methods*, 41(2), 295-308. doi: 10.3758/BRM.41.2.295
- McDaniel, M. A., Wildman, K. M., & Anderson, J. L. (2012). Using quizzes to enhance summative-assessment performance in a web-based class: An experimental study. *Journal of Applied Research in Memory and Cognition*, 1(1), 18-26. doi:10.1016/j.jarmac.2011.10.001
- McDonald, A. S. (2002). The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. *Computers & Education*. 39(3). 299-312. doi:10.1016/S0360-1315(02)00032-5



- McLaughlin, J. E., Griffin, L. M., Esserman, D. A., Davidson, C. A., Glatt, D. M., Roth, M. T., ..., & Mumper, R. J. (2013). Pharmacy student engagement, performance, and perception in a flipped satellite classroom. *American Journal of Pharmaceutical Education*, 77(3), 196. doi:10.5688/ajpe779196
- McNulty, J. A., Sonntag, B., & Sinacore, J. (2007). Test-taking behaviours on a multiple-choice exam are associated with performance on the exam and with learning style. *Journal of the International Association of Medical Science Educators*, 17(1), 52-57. [no doi available]
- Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114(3), 449-458. doi: 10.1037/0033-2909.114.3.449
- Niemiec, C. P., & Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom Applying self-determination theory to educational practice. *Theory and Research in Education*, 7(2), 133-144. doi:10.1177/1477878509104318
- Nikou, S., & Economides, A. A. (2013). Student achievement in paper, computer/web and mobile based assessment. Proceedings of the 6th Balkan Conference on Informatics (BCI), Greece
- Nonis, S. A. & Hudson, G. I. (2006). Academic performance of college students: influence of time spent studying and working. *Education for Business*, 81(1): 151-159. doi:10.3200/JOEB.81.3.151-159
- Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, 15(5), 625-632. doi:10.1007/s10459-010-9222-y
- Noyes J., Garland, K., & Robbins, L. (2004). Paper-based versus computer-based assessment: Is workload another test mode effect?. *British Journal of Educational Technology*, 35(1), 111–113. doi:10.1111/j.1467-8535.2004.00373.x
- Peterson, B. K., & Reider, B. P. (2002). Perceptions of computer-based testing: a focus on the CFM examination. *Journal of Accounting Education*, 20(4), 265-284. doi:10.1016/S0748-5751(02)00015-5
- Pierce, R., & Fox, J. (2012). Vodcasts and active-learning exercises in a "flipped classroom" model of a renal pharmacotherapy module. *American Journal of Pharmaceutical Education*, 76(10), 196. doi:10.5688/ajpe7610196
- Pintrich, P. R., Smith, D. A., García, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, 53(3), 801-813. doi:10.1177/0013164493053003024
- Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, 93(3), 223-231. doi: 10.1002/j.2168-9830.2004.tb00809.x
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Reckase, M. D., & Xu, J. R. (2014). The Evidence for a Subscore Structure in a Test of English Language Competency for English Language Learners. *Educational and Psychological Measurement*, 75(5), 805-825. doi:10.1177/0013164414554416
- Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, 95(2), 129-140. doi:10.1080/00223891.2012.725437

- 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. doi:10.1037/a0026838
- Ricketts, C., & Wilks, S. J. (2002). Improving student performance through computer-based assessment: Insights from recent research. *Assessment & Evaluation in Higher Education*, *27*(5), 475-479. doi:10.1080/0260293022000009348
- Rijksoverheid (2014). Toetsbesluit PO. Staatsblad. <https://www.rijksoverheid.nl/onderwerpen/toelating-middelbare-school/documenten/besluiten/2014/01/20/toetsbesluit-po>
- Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*, *17*(5), 1-25. doi:10.18637/jss.v017.i05
- Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, *130*(2), 261-288. doi:10.1037/0033-2909.130.2.261
- Robitzsch, R. (2016). sirt: Supplementary Item Response Theory Models. R package version 1.10-0. <http://CRAN.R-project.org/package=sirt>
- Roediger III, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181-210. doi:10.1111/j.1745-6916.2006.00012.x
- Roediger III, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, *17*(4), 382. doi:10.1037/a0026252
- Saint, D. A., Horton, D., Yool, A., & Elliott, A. (2015). A progressive assessment strategy improves student learning and perceived course quality in undergraduate physiology. *Advances in Physiology Education*, *39*(3), 218-222. doi:10.1152/advan.00004.2015
- Schneider, M. C., & Andrade, H. (2013). Lecturers' and Administrators' Use of Evidence of Student Learning to Take Action: Conclusions Drawn from a Special Issue on Formative Assessment. *Applied Measurement in Education*, *26*(3), 159-162. doi:10.1080/08957347.2013.793189
- Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, *15*(6), 657-680. doi:10.1007/BF01068419
- Schuman, H., Walsh, E., Olson, C., & Etheridge, B. (1985). Effort and reward: The assumption that college grades are affected by quantity of study. *Social Forces*, *63*(4), 945-966 doi:10.1093/sf/63.4.945
- Schuwirth, L. W., & Van der Vleuten, C. P. (2011). Programmatic assessment: from assessment of learning to assessment for learning. *Medical Lecturer*, *33*(6), 478-485. doi:10.3109/0142159X.2011.565828
- Sinharay, S., Wan, P., Choi, S. W., & Kim, D. I. (2015). Assessing Individual-Level Impact of Interruptions During Online Testing. *Journal of Educational Measurement*, *52*(1), 80-105. doi:10.1111/jedm.12064
- Sinharay, S., Wan, P., Whitaker, M., Kim, D. I., Zhang, L., & Choi, S. W. (2014). Determining the Overall Impact of Interruptions During Online Testing. *Journal of Educational Measurement*, *51*(4), 419-440. doi:10.1111/jedm.12052

- Sinharay, S. (2010). How often do subscores have added value? Results from operational and simulated data. *Journal of Educational Measurement*, 47(2), 150-174. doi:10.1111/j.1745-3984.2010.00106.x
- Sinharay, S., Puhan, G., & Haberman, S. J. (2011). An NCME instructional module on subscores. *Educational Measurement: Issues and Practice*, 30(3), 29-40. doi:10.1111/j.1745-3992.2011.00208.x
- Strayer, J. F. (2012). How learning in an inverted classroom influences cooperation, innovation and task orientation. *Learning Environments Research*, 15(2), 171-193. doi:10.1007/s10984-012-9108-4
- Street, S. E., Gilliland, K. O., McNeil, C., & Royal, K. (2015). The flipped classroom improved medical student performance and satisfaction in a pre-clinical physiology course. *Medical Science Educator*, 25(1), 35-43. doi:10.1007/s40670-014-0092-4
- Terzis, V., & Economides, A. A. (2011). The acceptance and use of computer based assessment. *Computers & Education*, 56(4), 1032-1044. doi:10.1016/j.compedu.2010.11.017
- Tomes, J. L., Wasylkiw, L., & Mockler, B. (2011). Studying for success: diaries of students' study behaviours. *Educational Research and Evaluation*, 17(1), 1-12. doi:10.1080/13803611.2011.563087
- Towns, M. H., & Robinson, W. R. (1993). Student use of test-wiseness strategies in solving multiple-choice chemistry examinations. *Journal of Research in Science Teaching*, 30(7), 709-722. doi:10.1002/tea.3660300709
- Tune, J. D., Sturek, M., & Basile, D. P. (2013). Flipped classroom model improves graduate student performance in cardiovascular, respiratory, and renal physiology. *Advances in Physiology Education*, 37(4), 316-320. doi:10.1152/advan.00091.2013
- Vaessen, B. E., van den Beemt, A., van de Watering, G., van Meeuwen, L. W., Lemmens, L., & den Brok, P. (2016). Students' perception of frequent assessments and its relation to motivation and grades in a statistics course: a pilot study. *Assessment & Evaluation in Higher Education*, 42(6): 872-886. doi:10.1080/02602938.2016.1204532
- Van der Drift, K. D. J., & Vos, P. (1987). Anatomie van een leeromgeving, een onderwijseconomische analyse of universitair onderwijs (Anatomy of a learning environment, an economic analysis of university education). Lisse, Netherlands: Swets and Zeitlinger.
- Vermunt, J. D., & Verloop, N. (1999). Congruence and friction between learning and teaching. *Learning and Instruction*, 9(3), 257-280. doi:10.1016/S0959-4752(98)00028-0
- Vermunt, J. D., & Vermetten, Y. J. (2004). Patterns in student learning: Relationships between learning strategies, conceptions of learning, and learning orientations. *Educational Psychology Review*, 16(4), 359-384. doi:10.1007/s10648-004-0005-y
- Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2008). Comparability of Computer-Based and Paper-and-Pencil Testing in K-12 Reading Assessments A Meta-Analysis of Testing Mode Effects. *Educational and Psychological Measurement*, 68(1), 5-24. doi:10.1177/0013164407305592
- Wei, X. & Haertel, E. (2011). The effect of ignoring classroom-level variance in estimating the generalizability of school mean scores. *Educational Measurement: Issues and Practice*, 30(1), 13-22. doi:10.1111/j.1745-3992.2010.00196.x

- West, S. G., Duan, N., Pequegnat, W., Gaist, P., Des Jarlais, D. C., Holtgrave, D., ... & Mullen, P. D. (2008). Alternatives to the Randomized Controlled Trial. *American Journal of Public Health, 98*(8), 1359–1366. <http://doi.org/10.2105/AJPH.2007.124446>
- Whitelock, D. (2009). Editorial: e-assessment: developing new dialogues for the digital age. *British Journal of Educational Technology, 40*(2), 199-202. doi:10.1111/j.1467-8535.2008.00932.x
- Wibowo, S., Grandhi, S., Chugh, R., & Sawir, E. (2016). A Pilot Study of an Electronic Exam System at an Australian University. *Journal of Educational Technology Systems, 45*(1), 5-33. doi:10.1177/0047239516646746
- William, D., & Black, P. (1996). Meanings and consequences: a basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal, 22*(5), 537-548. doi:10.1080/0141192960220502
- Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*(1), 3-17. doi:10.1207/s15326985ep2501\_2

## ICO Dissertation Series

In the ICO Dissertation Series dissertations are published of graduate students from faculties and institutes on educational research within the ICO Partner Universities: Eindhoven University of Technology, Leiden University, Maastricht University, Open University of the Netherlands, University of Amsterdam, University of Twente, Utrecht University, VU University Amsterdam, and Wageningen University, and formerly University of Groningen (until 2006), Radboud University Nijmegen (until 2004), and Tilburg University (until 2002). The University of Groningen, University of Antwerp, University of Ghent, and the Erasmus University Rotterdam have been 'ICO 'Network partner' in 2010 and 2011. From 2012 onwards, these ICO Network partners are full ICO partners, and from that period their dissertations will be added to this dissertation series.

List update January, 2018 (the list will be updated every year in January)

329. Wolff, C. (16-02-2016). *Revisiting 'withitness': Differences in teachers' representations, perceptions, and interpretations of classroom management*. Heerlen: Open University of the Netherlands.
330. Kok, E.M. (01-04-2016). *Developing visual expertise; from shades of grey to diagnostic reasoning in radiology*. Maastricht: Maastricht University.
331. De Beer, H.T. (11-05-2016). *Exploring Instantaneous Speed in Grade Five: A Design Research*. Eindhoven: Eindhoven University of Technology.
332. Ebbeler, J. (12-05-2016). *Implementing data use in schools: effects on the professional development of educators and the role of school leaders in data teams*. Enschede: University of Twente.
333. Draaijer, S. (10-06-2016). *Supporting Teachers in Higher Education in Designing Test Items*. Amsterdam: Vrije Universiteit Amsterdam.
334. Bos, L.T. (15-06-2016). *Moving Beyond Words. Supporting Text Processing Using a Situation Model approach*. Amsterdam: Vrije Universiteit Amsterdam.
335. Vrugte, J. ter (16-06-2016). *Serious support for serious gaming*. Enschede: University of Twente.
336. Kock, Z.D.Q.P. (23-06-2016). *Toward physics education in agreement with the nature of science: Grade 9 electricity as a case*. Eindhoven: Eindhoven University of Technology.
337. Trinh Ba, T. (28-6-2016) *Development of a course on integrating ICT into inquiry-based science education*. Amsterdam: Vrije Universiteit Amsterdam.
338. Gerken, M. (29-06-2016). *How do employees learn at work? Understanding informal learning from others in different workplaces*. Maastricht: Maastricht University.
339. Louws, M.L. (06-07-2016) *Professional learning: what teachers want to learn*. Leiden: Leiden University.
340. Geel, M.J.M. van, & Keuning T. (08-07-2016). *Implementation and Effects of a Schoolwide Data-Based Decision Making Intervention: a Large-Scale Study*. Enschede: University of Twente.
341. Bouwer, I.R., & Koster, M.P. (02-09-2016) *Bringing writing research into the classroom: The effectiveness of Tekster, a newly developed writing program for elementary students*. Utrecht: Utrecht University.
342. Reijners, P.B.G. (02-09-2016.) *Retrieval as a Cognitive and Metacognitive Study Technique to Learn from Expository Text*. Heerlen: Open University of the Netherlands.
343. Hubers, M.D. (08-09-2016). *Capacity building by data team members to sustain schools' data use*. Enschede: University of Twente.
344. Hsiao, Y.P. (23-09-2016). *Peer Support to Facilitate Knowledge Sharing on Complex Tasks*. Heerlen: Open University of the Netherlands.
345. Scheer, E.A. (23-09-2016). *Data-based decision making put to the test*. Enschede: University of Twente.
346. Bohle Carbonell, K. (28-9-2016). *May I ask you....? The influence of Individual, Dyadic, and Network Factors on the Emergence of Information in Exchange Teams*. Maastricht: Maastricht University.
347. Claessens, L.C.A. (30-09-2016). *Be on my side, I'll be on your side: Teachers' perceptions of teacher-student relationships*. Utrecht: Utrecht university.

348. Jansen in de Wal, J. (18-11-2016). *Secondary school teachers' motivation for professional learning*. Heerlen: Open University of the Netherlands.
349. Kock, W.D. de. (24-11-2016). *The effectiveness of hints during computer supported word problem solving*. Groningen: University of Groningen.
350. Oonk, C. (07-12-2016). *Learning and Teaching in the Regional Learning Environment: Enabling Students and Teachers to Cross Boundaries in Multi-Stakeholder Practices*. Wageningen: Wageningen University.
351. Beckers, J. (09-12-2016). *With a little help from my e-portfolio; supporting students' self directed learning in senior vocational education*. Maastricht: Maastricht University.
352. Osagie, E.R. (14-12-2016) *Learning and Corporate Social Responsibility. A study on the role of the learning organization, individual competencies, goal orientation and the learning climate in the CSR adaptation process*. Wageningen: Wageningen University.
353. Baggen, Y. (13-01-2017). *LLIGHT 'in' Europe - Lifelong Learning, Innovation, Growth and Human capital Tracks in Europe*. Wageningen: Wageningen University.
354. Wouters, A. (09-02-2017). *Effects of medical school selection. On the motivation of the student population and applicant pool*. Amsterdam: VU Medisch Centrum.
355. Baas, D.M. (01-05-2017). *Assessment for Learning: more than a tool*. Maastricht: Maastricht University.
356. Pennings, J.M. (04-05-2017). *Interpersonal dynamics in teacher-student interactions and relationships*. Utrecht: Utrecht University.
357. Lans, R.M. (18-05-2017). *Teacher evaluation through observation*. Groningen: University of Groningen.
358. Grohnert, T. (18-05-2017). *Judge/Fail/Learn; enabling auditors to make high-quality judgments by designing effective learning environments*. Maastricht: Maastricht University.
359. Brouwer, J. (22-05-2017). *Connecting, interacting and supporting. Social capital, peer network and cognitive perspectives on small group teaching*. Groningen: University of Groningen.
360. Van Lankveld, T.A.M. (20-06-2017). *Strengthening medical teachers' professional identity. Understanding identity development and the role of teacher communities and teaching courses*. Amsterdam: Vrije Universiteit Amsterdam.
361. Janssen, N. (23-06-2017). *Supporting teachers' technology integration in lesson plans*. Enschede: University of Twente.
362. Tuithof, J.I.G.M. (23-06-2017). *The characteristics of Dutch experienced history teachers' PCK in the context of a curriculum innovation*. Utrecht: Utrecht University.
363. Van Waes, S. (23-06-2017). *The ties that teach: Teaching networks in higher education*. Antwerp: University of Antwerp.
364. Evens, M. (30-06-2017). *Pedagogical content knowledge of French as a foreign language: Unraveling its development*. Leuven: KU Leuven.
365. Moses, I. (07-09-2017). *Student-teachers' commitment to teaching*. Leiden: Leiden University.
366. Wansink, B.G.J. (15-09-2017). *Between fact and interpretation. Teachers' beliefs and practices in interpretational history teaching*. Utrecht: Utrecht University.
367. Binkhorst, F. (20-10-2017). *Connecting the dots. Supporting the implementation of Teacher Design Teams*. Enschede: University of Twente.
368. Stoel, G.L. (14-11-2017). *Teaching towards historical expertise. Developing students' ability to reason causally in history*. Amsterdam: University of Amsterdam.
369. Van der Veen, M. (28-11-2017). *Dialogic classroom talk in early childhood education*. Amsterdam: Vrije Universiteit Amsterdam.
370. Frèrejean, J. (08-12-2017). *Instruction for information problem solving*. Heerlen: Open University of the Netherlands.
371. Rezende Da Cunha Junior, F. (19-12-2017). *Online groups in secondary education*. Amsterdam: Vrije Universiteit Amsterdam.
372. Van Dijk, A.M. (22-12-2017). *Learning together in mixed-ability elementary classrooms*. Enschede: University of Twente.

## **About the author**

In 2006 Anja completed High School at the Ukarumpa International School-Secondary Campus in Papua New Guinea. After this she studied *Educational Sciences* at the University of Utrecht from 2007-2010, including a 6 month exchange to the Australian National University in Canberra, Australia. From 2010 to 2013 she continued her studies at Utrecht University completing two research masters in *Educational Science: Learning in Interaction*, and *Methodology and Statistics for the Social and Behavioural Sciences*, while also working as a teaching and research assistant for Joop Hox, Edith de Leeuw, and Peter Lugtig. From 2013 to 2017 she worked on her PhD *Implementing Assessment Innovations in Higher Education*, at the University of Groningen with Rob Meijer, Roel Bosker, and Casper Albers. In 2017 she started working at the Free University of Amsterdam in the Methods and Applied Biostatistics group of the department of Health Sciences.



## Publications

- Albers, C. J., Boevé, A. J., & Meijer, R. R. (2015). A critique to Akdemir and Oguz (2008): Methodological and statistical issues to consider when conducting educational experiments. *Computers & Education*, *87*, 238-242. doi:10.1016/j.compedu.2015.07.001
- Boevé, A., Bronkhorst, L., Endedijk, M. D., & Meijer, P. (2015). Tackling methodological challenges to gain new insight into the complexity of student teacher learning in a dual teacher education program. In V. Donche, S. De Mayer, D. Gijbels, & H. van den Bergh (Eds.), *Methodological Challenges in Research on Student Learning*. (pp. 27-53). (Methodology and Statistics Series). AntwerpApeldoorn: Garant Publishers.
- Boevé, A. J., Meijer, R. R., Albers, C. J., Beetsma, Y., & Bosker, R. J. (2015). Introducing Computer-Based Testing in High-Stakes Exams in Higher Education: Results of a Field Experiment. *PLoS ONE*, *10*(12), doi:10.1371/journal.pone.0143616
- Boevé, A. J., Meijer, R. R., Bosker, R. J., Vugteveen, J., Hoekstra, R., & Albers, C. J. (2017) Implementing the flipped classroom: an exploration of study behaviour and student performance. *Higher Education*, *74*(6), 1015-1032. doi:10.1007/s10734-016-0104-y
- de Leeuw, E. D., Hox, J. J., & Boevé, A. (2016). Handling Do-Not-Know Answers: Exploring New Approaches in Online and Mixed-Mode Surveys. *Social Science Computer Review*, *34*(1), 116-132. doi:10.1177/0894439315573744
- Lugtig, P., Glasner, T., & Boevé, A. J. (2015). Reducing Underreports of Behaviors in Retrospective Surveys: The Effects of Three Different Strategies. *International Journal of Public Opinion Research*, *28*(4), 583-595. doi:10.1093/ijpor/edv032
- Meijer, R., Boevé, A., Tendeiro, J., Bosker, R., & Albers, C. (2017). The Use of Subscores in Exams in Higher Education: When is this Useful? *Frontiers in Psychology*, *8*, 305. doi:10.3389/fpsyg.2017.00305
- Meijer, R. R., Niessen, A. S. M., & Boevé, A. J. (2015). Rapporteren van subtest scores in de klinische praktijk: Oppassen met de interpretatie [A cautionary note on the use of subtest scores in clinical practice]. *De Psycholoog*, *50*, 35-41.
- Vugteveen, J., Boevé, A. & Hoekstra, R. (2018). The struggles of a field study: Studying the effectiveness of a flipped classroom. *SAGE Research Methods Cases*. doi:10.4135/9781526438188

## Submitted

- Boevé, A. J., Meijer, R. R., Beldhuis, H. J. A., Bosker, R. J., & Albers, C. J. *Natural Variation in Grades and its Implications for Assessing the Effectiveness of Educational Innovations in Higher Education*.
- Boevé, A. J., Albers, C. J., Bosker, R. J., Meijer, R. R., & Tendeiro, J. N. *Implementing Practice Tests in Psychology Education*.
- Evers, C., Dingemans, A., Junghans, A., & Boevé, A. J. *Feeling Bad Or Feeling Good, Does Emotion Affect Your Consumption Of Food? A Meta-Analysis Of The Experimental Evidence*



