

# Het gebruik van einddoeltoetsen bij aanvang van de studie

Citation for published version (APA):

Imbos, T. (1989). *Het gebruik van einddoeltoetsen bij aanvang van de studie*. [Doctoral Thesis, Maastricht University]. Rijksuniversiteit Limburg. <https://doi.org/10.26481/dis.19890101ti>

## Document status and date:

Published: 01/01/1989

## DOI:

[10.26481/dis.19890101ti](https://doi.org/10.26481/dis.19890101ti)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

## SUMMARY

This thesis discusses certain aspects of the Maastricht Progress Tests. Progress tests test the end objectives in the cognitive domain of medicine. The use of these tests in freshman years could be questioned. The research questions addressed in this thesis are:

- is it reasonable to assume that the progress tests include test items which can be answered by freshmen?

And if so,

- are there predictable differences and/or similarities between these test items and randomly selected test items?

A group of test items that can be answered by freshmen is called a Freshmen Test (FT). A group of test items that is randomly chosen is called a General Knowledge Test (GKT). Both types of tests were studied in the academic year 1983-1984 on four occasions. The differences and similarities between the tests are hypothesised. Three different kinds of hypotheses can be distinguished:

- hypotheses concerning the internal structure of the tests, i.e. differences in terms of the p-values, test scores, reliability.
- hypotheses concerning the external structure of the tests, i.e. differences in correlation patterns.
- hypotheses concerning the fit of the test data to the assumptions of psychometric models (CHAPTER 1).

To answer the first research question a panel of experts on the first year educational program were asked to judge 1029 test items. The inter rater reliability appeared to be acceptable. Between 19% and 24% of the test items corresponded to material handled in the first year program. The item selection method using content experts is compared to an empirical method using the observed p-values as a criterion. It could be shown that both methods were related. Furthermore the items that were selected by the content experts did not show a bias concerning difficult or easy test items. On the contrary the experts selected test items that had a wide range in difficulties. A conclusion is that the first research question may be answered positively (CHAPTER 2).

The differences concerning the internal structure between the FT and the GKT are studied. Hypotheses are tested in a sample of first year students, a sample of sixth

year students and a sample of physicians. Six out of nine hypotheses are fully confirmed by the results. Three hypotheses are only partly confirmed which means that the prediction was correct in some groups but incorrect in other groups. These results indicate that in the first year of medical school:

- FT test items are answered more than GKT test items.
- Ft test items are answered better than GKT test items.
- FT test items and GKT test items are equally reliable.
- FT test items indicate of educational effects more than GKT test items.
- The test scores on the FT test items differ significantly from the test scores on the GKT test items.

In the sample of sixth year students and the sample of physicians no differences between the FT and GKT test items regarding these aspects were found, except one: the test scores, which were significantly higher on the Freshmen Test, in both samples.

It was therefore concluded that as far as the internal differences between the tests are concerned the results indicate that the tests have construction validity. The Freshmen Test and the General Knowledge Test behave in most cases as was predicted in the hypotheses. (CHAPTER 3).

The hypotheses about the structure in the correlations between the four FT's and the four GKT's are formalised in two statistical models. In model I, one latent variable per occasion is specified. The covariance matrix of the latent variables is assumed to be generated by a simplex or first-order autoregressive model. The errors of measurement are assumed to be uncorrelated. In model II, the same structure as in model I is assumed, while, moreover, the errors of measurement of the same type of test are allowed to correlate. Two additional latent variables called test-specific factors (Jöreskog, 1970) are assumed to account for these correlations. It could be shown that model I fits the correlationmatrix in both the first year and the sixth year adequately. By using model II the fit of the model is significantly improved in the first year but not in the sixth year. Although, strictly speaking model, II is not needed to explain the correlations among the tests, the results indicate that some test-specific variance (variance due to the test being used) is present in the freshman years.

The predictive power of both the FT's and the GKT's appeared to be equal. The achievement of the cohort 1983-1984 could be predicted better for their second year than in the third year of medical school.

Finally the correlations between the FT's, the GKT's and other tests used to

measure medical knowledge during the first year of medical school are studied. As expected all the correlations were positive and most of them differed significantly from zero.

The hypothesised differences and similarities between the tests are strongly confirmed by the results. Indicating that, also from the external structure point of view, the tests demonstrate construct validity (**CHAPTER 4**).

In the next chapter hypotheses about the structure in the test results are formulated as psychometric models. These models are assumed to be minitheories for the construct medical knowledge. They are used to predict and test the structure in the responses of the students on the test items. Three models are described: the recently developed Progress Test model (PTmodel), the well-known Raschmodel and the Mokkenscalemodel. When applied in this context all models assume that the outcome of the interaction between a respondent and a test item is determined by two factors: the knowledge of the respondent and the difficulty of the test item. In the Raschmodel and in the Mokkenscalemodel these factors are parameters, while in the PTmodel these factors are assumed to be normally distributed random variables. (**CHAPTER 5**).

The testing results of the models showed that the fit of the PTmodel to all the test items is reasonable. Furthermore, at the level of the individual test items, about 50% of the items were Raschhomogeneous and about 34% of the items were congruent to the assumptions of the Mokkenscalemodel. These results indicate that about 80% of the individual test items have good measurement properties with monotonously increasing probabilities of a correct answer in the latent variable. Furthermore the analysis with the PTmodel showed that the FT test items had higher probabilities of a correct answer at all levels, except the highest, of the latent variable. The growth in knowledge among freshmen was highest on FT items, while the other years showed a higher growth in knowledge on GKT items. The Raschhomogeneous test items of both tests appeared on one occasion to be unidimensional. The Mokkenscale items of both tests were highly correlated. These results indicate that also at the level of individual test items, the assumption that the two tests measure the same trait cannot be rejected.

The analysis of the construct representation (Embretson, 1983) indicate that easier test items seemed to be related to so-called basic knowledge, whereas more difficult test items are related to subjects dealing with the practice of medicine. The conclusion is that, also at the psychometric level, the tests behave as was hypothesised (**CHAPTER 6**).

In the final chapter the results are discussed from four points of view:

- the meaning of the results for the use of progress tests in the freshman year.

- the results discussed from a cognitive point of view.
- the psychometric meaning of the results .
- the data-analytic findings in this study.

Although the results found in this study strongly indicate the construct validity of progress tests, the use of these tests for students with the low proficiency scores in the first year is not without problems. Suggestions to deal with the problem are discussed.

From the discussion of the results using the points of view mentioned, it is concluded that the results are congruent at the different theoretical levels. At each level new research questions emerged showing that the empirical domain of this subject is rich enough to justify further exploration.