

Moderation, a project in national prevocational education

Introduction

In the Dutch national exam system the candidates assignments are being marked by their own teacher guided by sophisticated compulsory marking schemes. The teachers marking results are checked and adjusted by a second marker, preferably, but not in every case, from another school. In this system, the lack of consistent checks by the Inspectorate provides a lot of freedom for schools in assessing their own pupils. This freedom is a potential and substantial threat to the quality of national exam standards.

To make teachers even more conscious of the importance of marking professionally and accurately Cito initiated a project to introduce moderation processes. This moderation process is mainly aimed at improving the professionalism of the teacher in his role of examiner. The moderation project started in 2007 as a pilot on a small scale, using video material that show students working on national vocational assignments. In the moderation sessions teachers are instructed how to assess these achievements which contain both process criteria as product criteria.

Dutch prevocational education

In the Dutch educational system at the age of 12 children have to choose to follow either pre-vocational secondary education or general secondary education. The fourth year concludes with a central (national) exam in general *and* vocational subjects. There are 36 different vocational subjects varying from care and welfare to masonry. The final year concludes with a central (national) exam in general and vocational subjects. The vocational subjects are assessed through an extensive national integrated practical and written exam in this final, fourth year. The length of this exam can extent to 920 minutes and is organized by the schools.

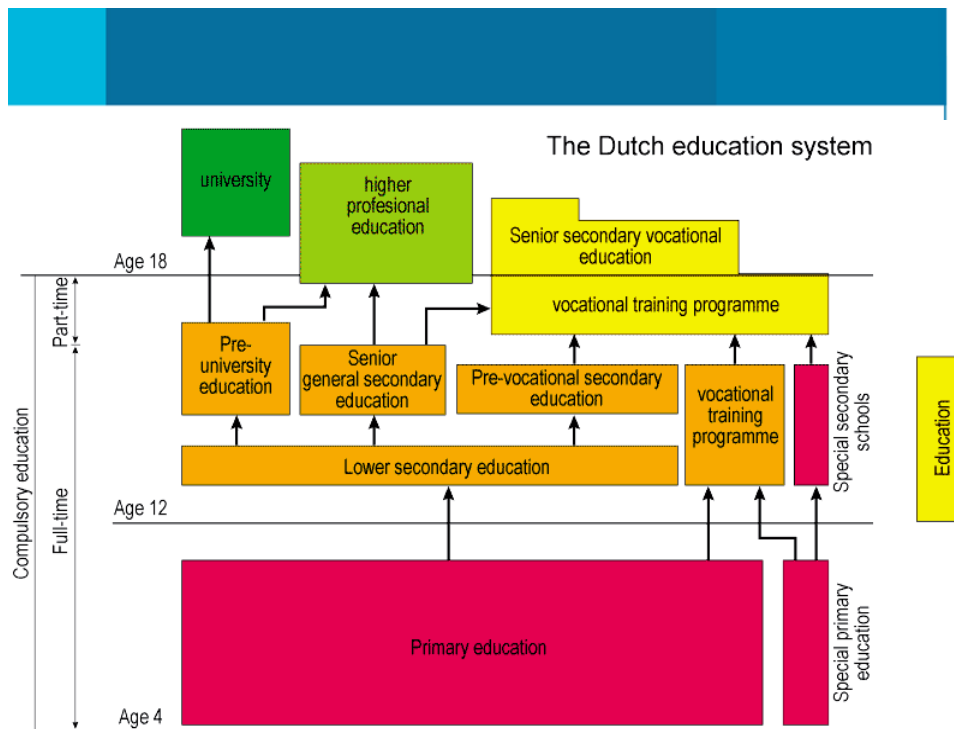
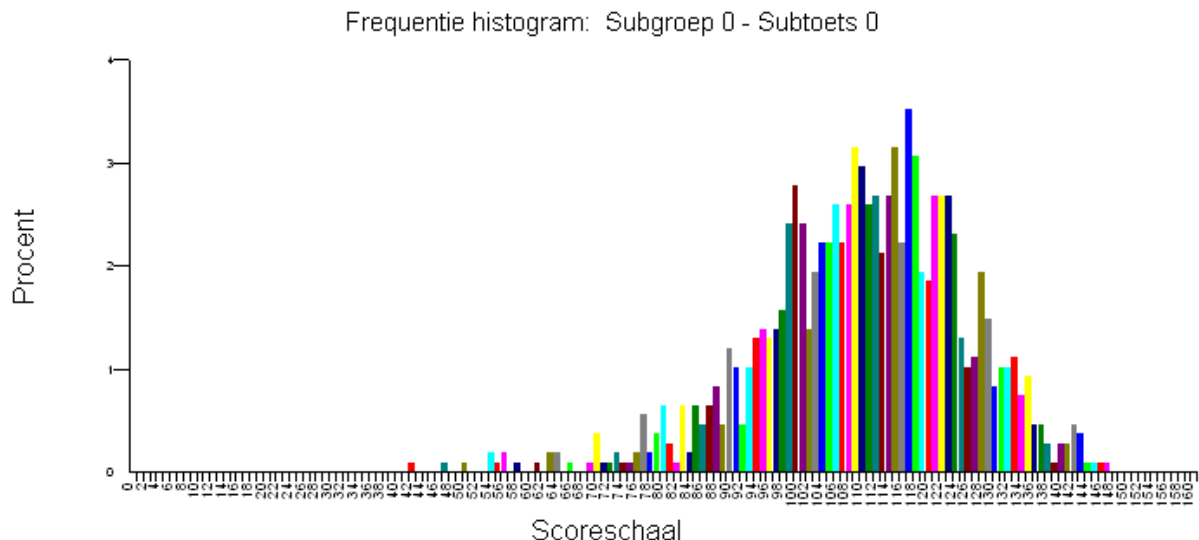


Figure 1 Dutch education system

A Problem and its interpretation

The problems experienced in this examination system come down to the following.

- The exams are too easy in general (according to the department of education) the percentage of candidates that fail was set at 15



SubGroepnummer	: 0	SubTest nummer	: 0
Aantal personen in toets	: 1079	Aantal geselecteerde items:	48
Minimum toetsscore	: 0	Maximum toetsscore	: 160
Gemiddelde toetsscore	: 110,11	Standaarddeviatie	: 15,52
Gemiddelde P-waarde	: 68,82	Standaardmeetfout	: 6,30
Coefficient Alpha	: 0,84	Standaardfout Alpha	: 0,01
Gemiddelde Rit	: 0,37		
GLB	: 0,91	Asymptotische GLB coef	: 0,90
Aantal items in GLB proc	: 48		

Figure 2 and 3, representative examples of data on pre-vocational national exams

The compact frequency distribution could be explained by the multidisciplinary character of the practical exams. Because this explanation is unsatisfactory to the politics Cito started to try to aim for more discrimination in the tests.

The first step was to improve the quality of the marking schemes by a number of means, using analytic and holistic polytomic scales combined with checklists consisting of dichotomous items.

This resulted in even more elaborate marking schemes, sometimes consisting of almost 200 criteria. One can state that sophistication knows its limits because the practical use of such a marking scheme started to be too complicated, especially in a workplace setting where 6 to 10 candidates work simultaneously while often only one assessor observes their performance.

The next step was to balance analytical marking schemes with user friendliness. This was achieved by eliminating all non-critical criteria and by reducing overlapping criteria (for instance do not judge all measures in an electric installation but take a random sample).

These new insights were also implemented successfully.

Nevertheless the difficulty stayed too low and the frequency distribution tables of the practical exams remained too compact. Cito's test experts often visited schools in the examination period and their observations indicated that examiners, not being fully independent, influenced the results of the exams. For teachers could tend to have a personal involvement with their students, especially with the low achievers.

Cito decided to look into this matter more precisely.

In 2007 Cito initiated a survey to investigate how raters applied our (compulsory) marking schemes. Seventeen test experts visited schools, observing a part of the examination, that sometimes takes several days, and interviewed the rater about how he or she had applied the rating criteria.

The following graph shows some of the results .

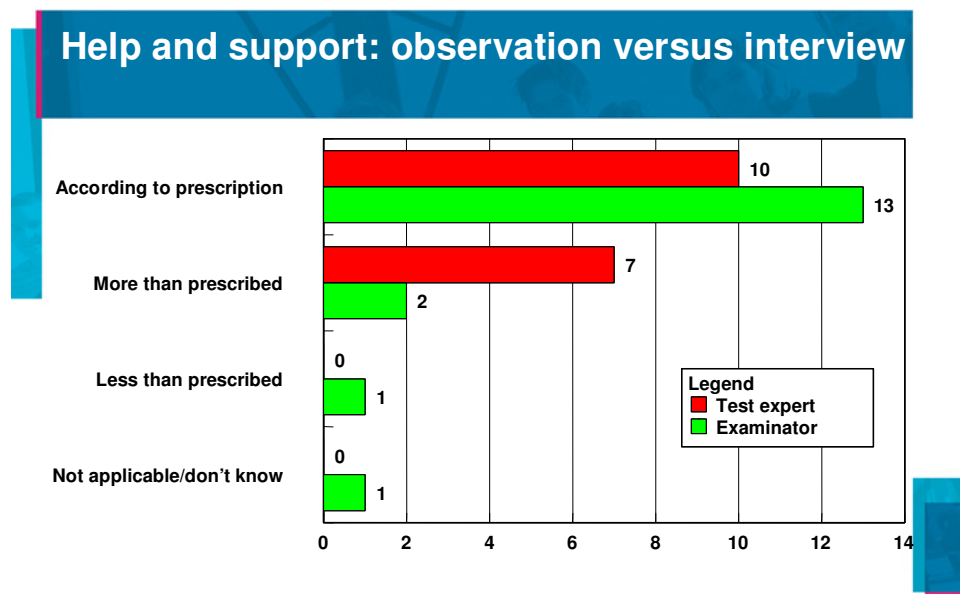


Figure 4 result of the survey

The assessment of the observing test expert is shown in red, the rater's self-assessment is represented in the green bars. The discrepancy between the perceptions of the observer and the rater is quite obvious. Apparently the rater is convinced that he has applied the schemes accurately in thirteen of the fourteen cases while the observer reports an accurate application in ten out of fourteen cases.

This is even more evident when we look at the second and the third bar from the top, in two cases the rater is aware that he has given more help than allowed while the observer counts seven cases of offering more help than allowed.

The following figure shows a similar situation. The top bar indicates that 9 raters state that they are never more lenient towards the candidate than allowed, while the test expert found evidence the rater assessed correctly without too much lenience in only 4 cases. The lower bars indicate some agreement between raters and observers concerning ratings being occasionally too lenient .

Lenient marking: observation versus interview

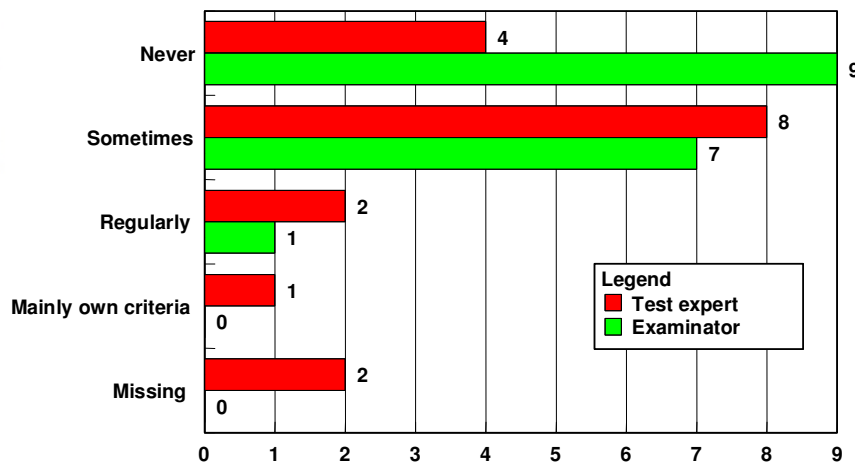


Figure 5 result of the survey

What can we conclude from this?

Raters have another perception about the application of the marking schemes than test experts. In a national examination setting this could lead to erosion of the syllabus and the national standards. When raters interpret and apply standards differently, candidates are assessed differently and could have advantage from having a more lenient assessor or in reverse have the disadvantage of having a harsh assessor.

What are the reasons for raters to assess more lenient than allowed?

We have identified a number of different motives.

- The rater is not only assessing the work of the students, but is unconsciously also judging his own teaching.
- The rater has a certain level of involvement with his students, especially with low achievers. This results in a dilemma in loyalty to either the student who could in the eyes of the examiner deserve to pass and the loyalty to the profession resulting in rating accurately.
- Teachers lack sufficient professionalism as examiner.

The project

To make teachers (even) more aware of their responsibility as an examiner, Cito developed a pilot in group moderation.

The goals of the pilot are :

- To clarify the evaluation criteria to the raters
- To enlarge the awareness of the examiners in their professionalism and their responsibilities in the national examinations
- To improve the quality of the assessment and the assessment techniques.

The Approach

Because towards the candidate lenience in assessment is most evident in practical work it seemed logical to improve the examiners agreement in this specific area. Two different subjects were selected, carpentry in the building techniques program and beauty treatment in the care & wellness program.

To train examiners in assessing practical work it was necessary to bring practical work into moderation sessions. Because assessment of the process is just as important in pre-vocational education as assessing the product it was clearly insufficient towards the candidate to bring only products into the moderation session. Especially, as in the case of beauty treatment where the product is usually on a models face. To resolve this dilemma students were filmed while working on old exam assignments. This video material was used in moderation sessions with 6 - 8 teachers / examiners that had to assess the students performance applying the marking scheme. This moderation session was chaired by a test expert and a subject expert.

The moderation process included several steps:

1. individual assessment of each rater
2. group moderation discussion aiming at raters agreement
3. individually questionnaires, investigating the differences in assessing the performance of the student before and after the moderation discussion
4. group discussion to evaluate the raters experiences with the moderation process and discuss possible learning effects.

The process

Immediately after starting the moderation project in 2007 the Dutch government introduced a new standard for secondary education implying that every student was entitled to at least 1040 hours of teaching a year. This had a severe impact on the project because of the decreasing availability of teachers. Suddenly they had problems getting permission to leave school for external activities, and in general, Dutch teachers, are, in contrary to some of their foreign colleagues, not very inclined to attend meetings in the weekends. Many teachers and students saw in this 1040 standard an analogy with poultry that has to stay inside during the threat of avian flu. For the project it appeared to be hard to find enough moderators, even when they were offered money, coverage of expenses and free meals.

A second problem was that the recruitment of moderators was organized through the professional associations, which made it more complicated to inform moderators accurately and assess their ability (e.g. if they really do teach the subject in question). The moderation sessions planned in april had to be cancelled because of a lack of moderators. The recruitment of moderators continued in order to be able to start a new series of sessions in late august 2008.

The results and follow up

The results are expected late August 2008 and are therefore not included in this paper. They will of course, be presented during the conference.

In the case of successful evaluation of the moderation pilot Cito intends to develop course material to professionalize examiners and to certify them. To work with certified examiners can make an attribution to the quality management of a school.

Rixt Zeelenberg works as a manager at Cito and is responsible for the production of the national exams in the technique department of the pre-vocational education. Technique consists of 17 subjects in three levels. Together 21 employees work on these exams.

Next to her management tasks she leads the project on moderation, works as a team member on a competention monitoring system for secondary education and is project leader in maintaining standards for national pre-vocational exams.

For more information contact: rixt.zeelenberg@cito.nl

Addendum

Questionnaire Evaluation moderation meeting

1. How exactly did you follow the marking scheme at the last exam?
 - a. Very accurate
 - b. accurate
 - c. More or less accurate
 - d. Not accurate
 - e. Not accurate at all

2. If you deviated from the marking scheme, what deviations did it concern?
 - a. I judged different aspects then the marking scheme indicated
 - b. I interpreted certain aspects in my own manner
 - c. I judged too strict
 - d. I judged too leniently
 - e. Others, being
 -
 -

3. If you judged a candidate leniently, for what reason did you do so?
 - a. I did not judge any candidate too lenient
 - b. To give the candidate the benefit of the doubt
 - c. Because the candidate benefits from "succes experience"
 - d. To not spoil the chances for the future for the candidate
 - e. To reward effort or good intentions
 - f. To take into account personal circumstances of the candidate, e.g. home situation
 - g. Because the subject or competence was not treated sufficiently in the classes
 - h. To take into account the manner in which the subject or competence was treated in class
 - i. To judge the candidate in the manner in which they are used to in class
 - j. Because I cannot pay attention to all candidates at the same time
 - k. Other reasons, being.
 -
 -
 - .
 -
 - .

4. To what extent do you agree with the following statements concerning the moderation session?

	Not agree at all	Not agree Totally agree	agree	
a. The general explanation about marking by the test expert was interesting	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. The general explanation about marking by the test expert was educational	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. The subject specific part treated by the test expert was interesting	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. The subject specific part treated by the test expert was educational	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e. The discussion with colleagues was interesting	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f. The discussion with colleagues was educational	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
g. The session made me more aware of the value of an objective assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
h. In this session I learned more about my subject	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
i. This session made me a better test marker	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
j. At a next practical exam I will mark the exams in a different manner	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

5. Would you advise your colleagues to take part in a moderation session by Cito?
 - a. Yes
 - b. No
 - c. Don't know

6. Would you be prepared to facilitate a moderation session for you colleagues at school?
 - a. Yes
 - b. No
 - c. Don't know

7. Cito is considering to develop course material to enable schools to organize their own moderation sessions. Do you think you could organize such a moderation session without external assistance (with the provided course material).
 - a. Yes
 - b. No
 - c. Don't know

8. Cito is considering schooling several employees in supervising and supporting moderation sessions in schools. Cito would offer this supervision and support at cost price. Do you think your school would be prepared to pay for this external assistance?
 - a. Yes
 - b. No
 - c. Don't know

9. Please rate the moderation session from 1 to 10?