DOCUMENT RESUME

ED 362 041 FL 021 505

AUTHOR	Krieken, Robert van
TITLE	Equating National Exams in Foreign Language Reading Comprehension.
INSTITUTION	Centraal Inst. voor Toetsonwikkeling, Arnhem (Netherlands).
PUB DATE	9 Jul 93
NOTE	14p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Foreign Countries; *Reading Comprehension; *Scores; Secondary Education; Second Language Instruction; *Second Language Learning; Test Construction; Tests
IDENTIFIERS	*Netherlands

ABSTRACT

The procedures for the construction of the Dutch national exams of reading comprehension and for setting cut-off scores have remained roughly unchanged over 20 years. Construction procedures are characterized by thorough screening rather than pretesting; cut-off scores are influenced by procedures and percentages fails rather than by equating. Two studies were carried out to demonstrate the necessity and feasibility of equating procedures using IRT methodology. The first study equated exams from 1984 to 1990 with the same old exam using section post-equating and producing differences in mean difficulty. Traditional estimates and IRT estimates lead to the same cut-off scores. The second study was part of a large scale project the Inspectorate had commissioned, investigating the equivalence of cut-off scores in 17 subjects as well as differences in populations using teachers' estimates as well as empirical data. Data were scaled using IRT methodology, producing estimates of the mean score candidates in 1991 would have made on previous exams. The first study demonstrated differences between exams and showed that estimates were robust. The second study showed that teachers' estimates were consistent, but correlate only moderately with pupils' results. (Author/JL)

Cito Instituut voor Toetsontwikkeling

25 jaar Cito

Nieuwe Oeverstraat 65
Telefoon (085) 52 11 11
Telefax (085) 52 13 56
Postbus 1034
6801 MG Arnhem
Postrekening 1745300
ABN/Amro 43 82 18 558

EQUATING NATIONAL EXAMS IN FOREIGN LANGUAGE READING COMPREHENSION

Robert van Krieken

Kenmerk
LTRC'93

Datum
9 juli 1993

Abstract

The procedures for the construction of the Dutch national exams of reading comprehension and for setting cut-off scores have remained roughly unchanged for over twenty years. Construction procedures are characterized by thorough screening rather than pretesting; cut-off scores are influenced by procedures and percentages fails rather than by equating. Cito carried out two studies to demonstrate the necessity and feasibility of equating procedures using IRT methodology.

The first study equated exams from 1984 till 1990 with the same old exam using section post-equating and producing differences in mean difficulty. Traditional estimates and IRT estimates lead to the same cut-off scores. The second study was part of a large scale project the Inspectorate had commissioned, investigating the equivalence of cut-off scores in 17 subjects as well as differences in populations using teachers' estimates as well as empirical data. Data were scaled using IRT methodology, producing estimates of the mean score candidates in 1991 would have got on previous exams and comparing these with the actual mean scores of previous populations.

The first study demonstrated differences between exams and showed that the estimates were robust. The second study showed that teachers' estimates were consistent, but correlate only moderately with pupils' results. Here again cut-off scores differed. About one out of every six previous cut-off scores turned out to be not equivalent to the most recent one. The population means varied too, so the distribution should not be taken (as it is) as a starting point for setting the norm. Acting upon the outcomes of the second study, the State Secretary for Education and Science has provided funds for introducing and maintaining equating as a standard procedure in central exams.

keywords: central examinations, equating, teachers' estimations, population means.

1 BACKGROUND AND RATIONALE

National exams in Holland consist of two parts, a school internal part which is constructed, set and marked by the individual schoolteacher, and a central part which is constructed by Cito, but set and marked according to procedures laid down by the committee for central exams in secondary education (CEVO).

For the foreign languages the central exam tests only reading comprehension by means of 50 multiple choice items. The mother tongue on the other hand, is centrally examined on two aspects, on reading comprehension by means of a mixture of multiple choice and open-ended questions, and on writing by means of either an essay or a number of functional assignments. Characteristic for the procedures concerning central examinations is that every year for each exam the construction cycle starts all over again, that every year the cut-off score is set anew, within a narrow range, and that in general the CEVO prefers screening to pretesting.

Against this background it isn't surprising that CITO has investigated the possibilities of relating the exams to each other providing information on their relative difficulty in order to ascertain consistency of cut-off scores.
To this end a series of equatings have been carried out, resulting in the advice to introduce section post equating as a standard part of the procedure. This first study offered Cito the possibility of applying IRT based equating programs (Glas 1988) in real practice and to compare the outcomes with classical equating using multiple matrix sampling (Shoemaker 1973).

In Parliament there has been some doubt as to whether the increasing numbers of pupils opting for higher forms of education have not been accommodated by declining standards. This led to a major research project from the Inspectorate, conducted by Cito, into the level of cut-off scores and into the ability of candidates over the period 1981-1991. In this second study it was possible to use Rasch homogeneous subscales to express the difficulties of all exams on one common scale (Glas 1989).

2 FIRST STUDY

design and methods

In the first study reading comprehension exams have been equated for several years up to 1990, starting with a pilot project in 1984 (Sanders & Goldebeld 1985) which led to an official experiment in 1988 and 1989. Here only the results for German, French, English and Dutch (L1) reading comprehension for '87-'90 are reported. In all cases the equating consisted of comparing the same old exam (called the reference exam) with the current one. In most cases both exams were split up into sections, and booklets were formed consisting of one section of each (see figure 1). Only the Dutch reading comprehension exams couldn't be divided into sections as they contained questions comparing two or more texts, so these exams were set as a whole, assuming that pupils who took one exam were equivalent to those who took the other. In all cases all different versions were distributed within each class.

The exams were meant for pupils completing the fourth and last year of the lowest type of general and vocational secondary education. For reasons of confidentiality, equating could not take place before the actual exam. Therefore the booklets were presented to pupils in the third year of a higher type of education in the week immediately following the actual exam. The pupils performed slightly differently from the actual exam population. As the purpose of the equation was a comparison within these non-equivalent groups this difference was of no statistical importance.

book let	reference exam 50 items sections 1 to 5	current exam 50 items sections 1 to 5
1	1	3
2	2	4
3	3	5
4	4	1
5	5	2

Figuur 1: distribution of sections from two exams over booklets

Mean scores for both exams were estimated using MULTIMAX (Shoemaker 1973) for multiple choice exams and VARCOM/GSS (Tormäkängas) for the Dutch exam containing polytome questions. In 1989 and 1990 the pairs of multiple choice exams were also scaled on one common Rasch scale assuming unidimensionality. The equivalent cut-off score for the current exam was estimated in the simplest possible way, by using only the difference between the means. So, if the mean score for the current exam proved to be two points lower than that for the reference exam, the advice would be to set the cut-off score two points lower than that for the reference exam. This seemed to be most transparent for subject specialists. In fact, differences in variance between exams were slight and hardly affected the choice of cut-off scores.

To gather more information about the robustness of this method of estimating equivalent cut-off scores, the equating was replicated in three different ways. Out of eight replications, only one gave a minimally different result. It can therefore be concluded that they are sufficiently robust. A first replication, for the German exam of 1988, made use of pupils from the educational type the exam was meant for, instead of a non-equivalent group, one month before the exam, and produced Rasch analyses. This yielded the same results. Secondly, Rasch analyses were used in reanalysing the 1989 data for the three foreign languages. This too led to the same estimates as resulted from the classic equating. In a third replication, also for each of the three foreign languages, 1000 pupils from the authentic exam population with a score above guessing level were added to the design. Then all pupils were divided into three score groups in each of which both exams were equated using the Rasch model. Here there was one case where the result was different from before, though with only one point (Glas 1989: 106-109).

Results

From figure 2 it is clear that exams do show marked differences with the reference exam and from one year to another.

In 5 out of fourteen exams the committee did not apply the advised equivalent cut-off score: 3 times they set a more lenient, 2 times a more severe cut-off score (see appendix 1 for more details). It is important to note that this committee is fully authorized to do that. The reasons for following another course were diverse. In three cases there were formal reasons: the equivalent cut-off scores would have fallen outside the allowed band or the difference with the usual cut-off score was too slight to change it. In two other cases

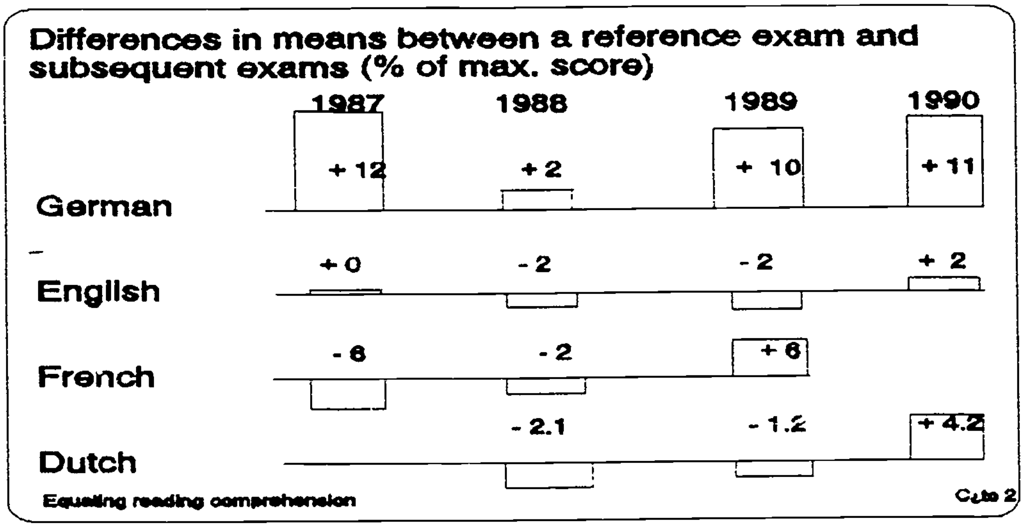


Figure 2: Differences between current exams and a reference exam
+ : current exam is easier than reference exam (higher cut-off score)
- : current exam is more difficult (lower cut-off score)

arguments other than equivalence prevailed: in 1988 the advised equivalent cut-off score for English was thought to be too lenient for the assumedly large influx of pupils from vocational education, in 1989 the committee decided that the French reference exam and its cut-off score had become too difficult for this level.

From discussions with the committees responsible for the exams it appeared that they found it difficult to understand how the equivalent cut-off scores were arrived at. Especially the use of within-group comparisons in non-equivalent groups were found hard to accept.

2 SECOND STUDY

design and methods
The second study consisted of two parts. In the first part exams were distributed over teachers with experience in exam classes. They were asked to rate the difficulty of all questions from a number of exams as well as the difficulty of the integral exams. In the second part, sections of five central exams were set for pupils preparing themselves for their exams. By comparing the teachers' judgements with the pupils' results, their usefulness as predictors of difficulty could be evaluated. Figure 3 gives a survey of the exams used in the second study.

LANGUAGE AND LEVEL[1]	81	82	83	84	85	86	87	88	89	90	91
German D		X		X		X		X			X
English C	X		X		X			X			X
English D	X		X		X			X			X
English HAVO	X		X			X			X	X	
English VWO				X	X	X		X			X
French VWO	X			X		X			X		X
Dutch C (L1)						X	X	X		X	X

Figure 3: Selection of exam years in the design

[1] After four years of general secondary (MAVO) or preparatory vocational education (VBO, formerly called LBO) pupils may choose C or the higher D-level; HAVO-level is attained after five years of general secondary education preparing for higher vocational training and VWO-level is attained after six years of general secondary education preparing for academic study.

2.1 SECOND STUDY, PART ONE: TEACHERS' ESTIMATIONS

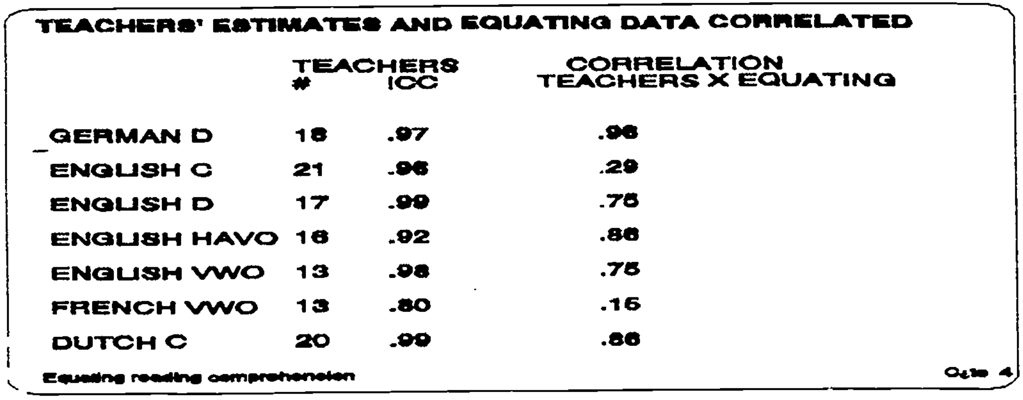
In this part of the study groups of 13 to 21 teachers participated who had extensive experience in exam classes, and sometimes experience in exam construction as well. They all had been using old exams in the classroom, but were advised not to look up any information about their difficulty. They were asked to rate the difficulty of each item from the exams in the design on a scale. The instruction they were provided with contained a scale with eight categories, each of them illustrated with one or more items. These examples were chosen from another exam and had difficulty indices at about the middle of each class as in table 1.

Table 1: difficulty categories with their corresponding difficulty level.

category	difficulty level (% of maximum score)
1	appr. 30
2	appr. 45
3	appr. 55
4	appr. 65
5	appr. 70
6	appr. 75
7	appr. 85
8	appr. 95

The teachers' classifications have been transformed to mean difficulty indices for total exams by taking the middle of the category as the difficulty level and computing the mean over all items in the same exam. In the Dutch exams, reading comprehension questions were first weighted according to their maximum score. Then interrater reliabilities were computed as well as correlations between the mean over all teachers' estimations and the mean pupils' scores (as estimated in the second part of study 2), and the differences between both means.

Results of teachers' estimations



TEACHERS' ESTIMATES AND EQUATING DATA CORRELATED

	TEACHERS #	ICC	CORRELATION TEACHERS X EQUATING
GERMAN D	16	.97	.96
ENGLISH C	21	.96	.29
ENGLISH D	17	.99	.75
ENGLISH HAVO	16	.92	.86
ENGLISH VWO	13	.98	.75
FRENCH VWO	13	.80	.15
DUTCH C	20	.99	.86

Equating reading comprehension

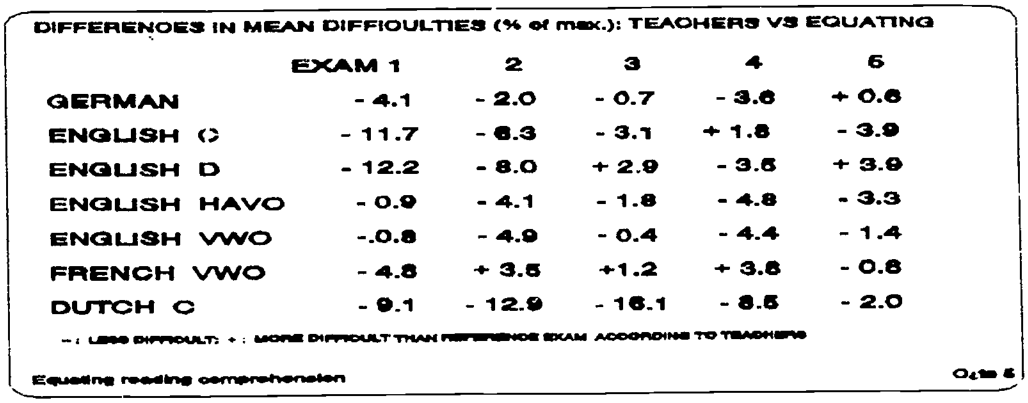
Dia 4

Figure 4: Teachers' estimates and equating data

Figure 4 offers a survey of the interrater reliabilities and also of the correlation between the teachers' estimations of the difficulty and the difficulty as computed in the second part. The reliabilities are intra-class correlations for the whole group. They reflect agreement in order, not counting systematic differences in means. All groups were found to be highly or very highly reliable. This means that other estimations by the same or comparable groups would come to almost precisely the same estimations.

The correlations between the teachers' estimations and the pupils' means clearly vary. For two groups they are very low, for two others moderate, and only for three of the seven groups are they really high.

The usefulness of the estimations is illustrated by figure 5, which shows the difference between the mean teachers' estimates and the pupils' mean scores in percentages of the maximum score. Teachers are clearly optimistic. But not constantly in the same degree. Differences of 2% amount to one score point in the foreign language exams. Estimates that are exactly right one year but two points wrong the following year are not sufficiently precise to be useful.



DIFFERENCES IN MEAN DIFFICULTIES (% of max.): TEACHERS VS EQUATING

	EXAM 1	2	3	4	5
GERMAN	- 4.1	- 2.0	- 0.7	- 3.6	+ 0.6
ENGLISH C	- 11.7	- 6.3	- 3.1	+ 1.8	- 3.9
ENGLISH D	- 12.2	- 8.0	+ 2.9	- 3.5	+ 3.9
ENGLISH HAVO	- 0.9	- 4.1	- 1.8	- 4.8	- 3.3
ENGLISH VWO	-.0.8	- 4.9	- 0.4	- 4.4	- 1.4
FRENCH VWO	- 4.8	+ 3.5	+1.2	+ 3.6	- 0.8
DUTCH C	- 9.1	- 12.9	- 16.1	- 8.5	- 2.0

- : LESS DIFFICULT; + : MORE DIFFICULT THAN REFERENCE EXAM ACCORDING TO TEACHERS

Equating reading comprehension

Dia 5

Figure 5: Differences in means between teachers' and equating data

CONCLUSIONS PART ONE

Teachers estimated the difficulty of exams with a high enough degree of agreement to be called precise. Compared with data from equating however, they underestimated the difficulty and they did this inconsistently. Their precise estimations were not correct. It is to be remembered that the study used old exams which were known to the teachers and which they used to read with their classes. If teachers were to participate in a procedure to equate exams that are quite new to them this would hardly produce better results. The result made it abundantly clear that teachers' estimations generally cannot replace equating using data collected among pupils.

2.2 SECOND STUDY, PART TWO: DATA COLLECTED AMONG PUPILS

Design and method

The same exams that were judged by the teachers were also set to pupils reading for their finals. The administration took place two months before the actual exams. At the beginning of the year the teachers had been instructed not to discuss the five exams under study.

The total amount of texts and items was distributed over a large number of booklets in such a way that each booklet could be answered within two consecutive school periods and that all booklets were connected by overlapping parts (see the example in appendix 3).

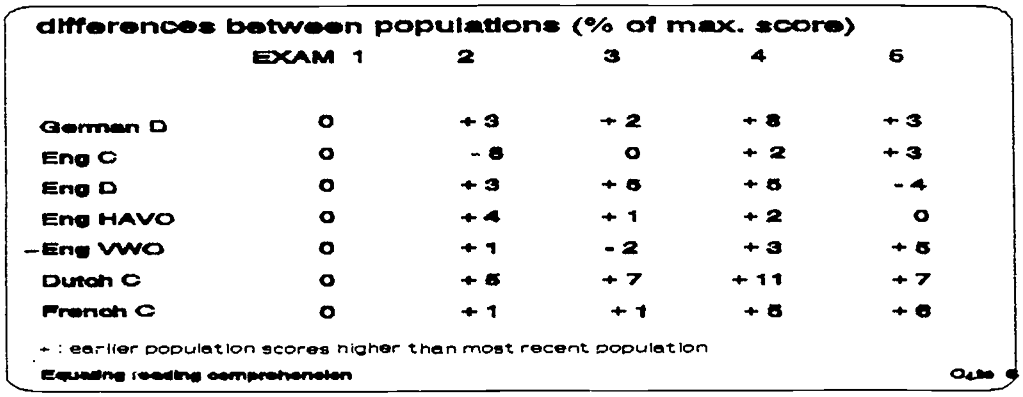
On the basis of the results first the difficulties of all complete exams were estimated on one common scale. This was done by creating as many subscales as were necessary to fit the Rasch model and then combining them to a common scale (Glas 1989). In a second step these difficulties were used together with the known score distributions of about 1000 authentic candidates on the most recent exam to estimate which scores these recent candidates would have got on each previous exam. This meant that two comparisons could be made: first of all the percentage fails that the most recent population would have showed had they taken an old exam, could be compared with the percentage fails within that year's population. This would answer the question whether there has been a decline in ability among the candidates. Secondly, equivalent cut-off scores could be estimated and compared with the actual cut-off scores. Estimating equivalent cut-off scores was done by equipercentile method: taking the percentage fails in the most recent population as a point of reference and looking up the score in every older exam at which this same population would have produced almost the same percentage of fails. This would answer the question whether there had been a decline in norms.

Results

Detailed results of the equation are shown in appendix 4. Here we present only differences between the actual and the equivalent cut-off scores and between populations. The differences between the most recent population (1991, except for English HAVO where the 1990 population was the most recent) and candidates from previous years is shown by figure 6.

Although most previous populations appear to have done better than the most recent population would have done, there is no clearly discernable decline in the sense that each new population scores lower than the former one.

Figure 7 shows the differences between the actual cut-off scores and those that have been estimated to be equivalent. Not all differences are meaningful, of course. Some might be due to statistical error. As the usual



differences between populations (% of max. score)

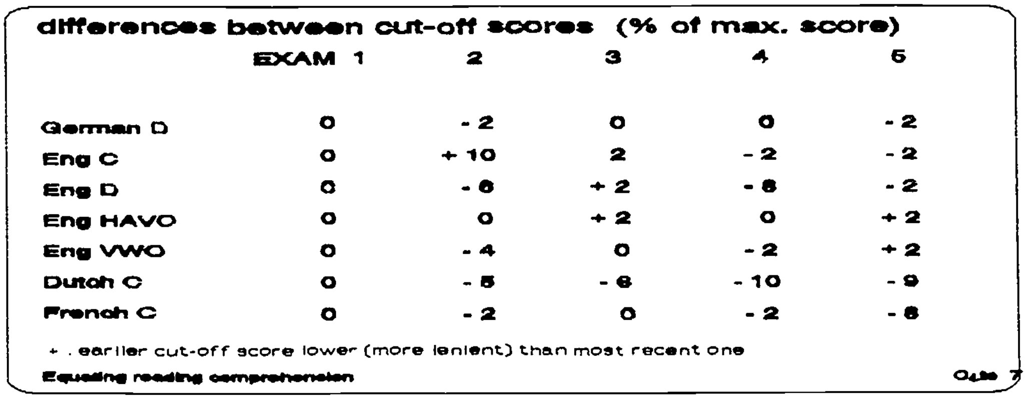
	EXAM 1	2	3	4	6
German D	0	+ 3	+ 2	+ 8	+ 3
Eng C	0	- 8	0	+ 2	+ 3
Eng D	0	+ 3	+ 5	+ 5	- 4
Eng HAVO	0	+ 4	+ 1	+ 2	0
–Eng VWO	0	+ 1	- 2	+ 3	+ 5
Dutch C	0	+ 5	+ 7	+ 11	+ 7
French C	0	+ 1	+ 1	+ 5	+ 6

+ : earlier population scores higher than most recent population

Equating reading comprehension Oslo 6

Figure 6: differences between previous populations and the most recent one.

standard error of an exam is about 6% of its total score, it seems reasonable to take only differences of at least 7% into consideration. There are but 5 of these major differences in 28 exams, randomly distributed over subjects and years. It is interesting to note, though, that in the only case where the actual cut-off score has been much more lenient than that of 1991, the population had much lower scores than the 1991 population would have had. In the other cases lower, i.e. more severe a:tual cut-off scores correspond with higher performances of the actual previous populations. As a matter of fact, there is a correlation of .49 between diferences in population and those in cut-off scores. This clearly suggests a tendency to use the score distribution and set the cut-off score at an acceptable percentage fails without taking into consideration that the whole population might perform better or worse than before.



differences between cut-off scores (% of max. score)

	EXAM 1	2	3	4	5
German D	0	- 2	0	0	- 2
Eng C	0	+ 10	2	- 2	- 2
Eng D	0	- 6	+ 2	- 8	- 2
Eng HAVO	0	0	+ 2	0	+ 2
Eng VWO	0	- 4	0	- 2	+ 2
Dutch C	0	- 5	- 6	- 10	- 9
French C	0	- 2	0	- 2	- 8

+ . earlier cut-off score lower (more lenient) than most recent one

Equating reading comprehension Oslo 7

Figure 7: differences between equivalent and actual cut-off scores

3 IMPLICATIONS

The first study showed that exams clearly don't have the same difficulty. When confronted with the difference between the current exam and an exemplary previous one the committees concerned have in most cases acted upon this information and set the advised equivalent cut-off score. The use of non-equivalent groups however, and the statistical methods used have not been understood or accepted.

The second study has confirmed the intuition of the teachers most directly involved in test construction, that they cannot be expected -as they are- to predict the difficulty of a new exam with sufficient precision. This would be a minor problem if we could assume that the ability level of the candidates remained constant over the years. Then fluctuations in the mean scores between the current exam and previous ones that show up after the exams are analysed could be seen as pure indications of the difficulty of the exams. This assumption turned out to be false too. Populations do vary (but, contrary to expectations, no decline could be demonstrated).

The second study also showed that one out of every six cut-off scores is clearly not equivalent with that of 1991 (not counting differences among the previous exams themselves). In the case of Dutch reading comprehension this is the result of a well-considered decision to lower the norm. In other cases the non-equivalent cut-off scores could be caused by the absence of an equating procedure. The first (exploratory) study providing this only covered part of the period 1981-1991, and didn't cover 1991 at all being discontinued immediately after the exams of 1990. The future looks brighter, however. Acting upon the outcomes of the second study, the State Secretary for Education and Science has promised to fund the introduction of equating as a standard procedure for central examinations.

Note I'd like to thank Kees Glas and Maarten Groot for the way they have taken care of the psychometric part of the research, Michel Zwarts for his advice concerning the second study, Noud van Zuijlen for his many quick reactions to several drafts and the other people from the examinations department for the trouble they have taken to understand what we have been doing and explain the use of it to their committees.

REFERENCES

Brennan, Robert L. & Michael J. Kolen. (1987) Some practical issues in equating. *Applied Psychological Measurement* 11, 3 p. 279-290.

Glas, C.A.W. (1989). *Contributions to Estimating and Testing Rasch models*. Proefschrift, Universiteit Twente.

Krieken, R. van. (1990) *Equating in the Dutch centralized examinations*. paper IAEA Maastricht.

Krieken, R. van. (1993) *Ontwikkeling van examennormen. Verslag van een onderzoek t.b.v. de inspectie*. Cito, Arnhem.

Masters, G.N. (1982). A Rasch model for partial credit scoring. **Psychometrika, 47**, 149-174.

Rasch, G. (1960). *Probablistic models for some intelligence and attainment tests*. Danish Institute for Educational Research, Kopenhagen.

Sanders, P. & P. Goldebeld. (1985) *Het pre-equivaleren van examens*. ORD paper

Shoemaker, D.M. (1973) *Principles and procedures of Multiple Matrix Sampling*. Cambridge, Bollinger

APPENDIX 1: EQUATING DATA FIRST STUDY

LANGUAGE	YEAR	PUPILS	MEAN SCORE		CUT-OFF POINT		
			REFERENCE (max. 50)	CURRENT (max. 50)	REFERENCE	EQUIVALENT	ACTUAL
GERMAN D							
_	1987	539	30.4	36.4	27.5	33.5	31.5
	1988	890	30.1	31.4	27.5	28.5	28.5
	1989	579	31.1	35.9	27.5	32.5	32.5
	1990	1878	30.0	35.2	27.5	32.5	32.5
ENGLISH D	1987	775	34.8	34.9	29.5	29.5	29.5
	1988	522	33.7	33.1	29.5	28.5	29.5
	1989	874	37.5	36.5	29.5	28.5	28.5
	1990	1045	37.2	37.9	29.5	30.5	30.5
FRENCH C	1987	584	37.2	34.5	25.5	22.5	24.5
	1988	1111	37.3	36.3	25.5	24.5	24.5
	1989	644	36.1	38.9	25.5	28.5	26.5
DUTCH D	1988	257	30.6	28.5	27.5	25.5	24.5
	1989	445	30.6	29.4	27.5	26.5	26.5
	1990	1103	63.9%[1]	68%	55%	50%	50%

[1] From 1990 on the maximum score has been raised from 50 to 90 points.

APPENDIX 2: SECOND STUDY, PART ONE, TEACHERS' ESTIMATIONS

A	B	C	D	E .	F	G
Lang	'Yr	-- Teachers'		--	pupils'	diff
		mean	#	Sd	mean	F-C
	%	%				
GERD	82	65.14	18	5.64	61	-4.14
GERD	84	69.01	18	7.00	67	-2.01
GERD	86	67.73	18	5.29	67	-0.73
GERD	88	66.60	18	6.25	63	-3.60
GERD	91	69.39	18	6.87	70	0.61
ENGC	81	69.74	21	4.51	58	-11.74
ENGC	83	72.26	20	5.92	66	-6.26
ENGC	85	70.90	21	5.08	74	3.10
ENGC	88	68.76	21	5.15	67	-1.76
ENGC	91	71.89	20	5.23	68	-3.89
ENGD	81	66.19	17	5.71	54	-12.19
ENGD	83	72.04	16	5.26	64	-8.04
ENGD	85	72.06	17	4.58	75	2.94
ENGD	88	70.47	17	4.17	67	-3.47
ENGD	91	70.12	16	5.[illegible]	74	3.88
ENGH	81	68.94	16	4.74	68	-0.94
ENGH	83	69.13	16	4.90	65	-4.13
ENGH	86	67.84	16	6.42	66	-1.84
ENGH	89	65.84	16	5.11	61	-4.84
ENGH	90	66.34	16	4.78	63	-3.34
ENGV	84	75.85	13	5.77	75	-0.85
ENGV	85	72.88	13	6.94	68	-4.88
ENGV	86	70.44	12	6.24	70	-0.44
ENGV	88	73.35	13	5.19	69	-4.35
ENGV	91	75.38	13	6.90	74	-1.38
FREV	81	67.76	10	6.15	63	-4.76
FREV	84	68.48	13	5.59	72	3.52
FREV	86	66.85	13	3.69	68	1.15
FREV	89	67.22	13	6.31	71	3.78
FREV	91	69.82	13	6.44	69	-0.82
DUTC	86	62.13	20	4.32	53	-9.13
DUTC	87	65.93	20	4.92	53	-12.93
DUTC	88	61.10	20	5.82	45	-16.10
DUTC	90	70.53	20	6.89	62	-8.53
DUTC	91	68.01	20	6.22	66	-2.01

APPENDIX 3: DISTRIBUTION OF EXAMS OVER BOOKLETS, GERMAN-D

BOOKLETS	ITEMS FROM EXAM YEAR				
	1981	1983	1985	1988	1991
1	1-8, 20-30	11-30			
2		11-30	1-10, 33-41		
3			1-10, 33-41	30-50	
4				30-50	11-31
5	9-19, 31-40				11-31
6	9-19, 31-40	31-50			
7		31-50	22-32	11-19	
8			22-32	11-19	32-50
9	41-50	1-10			32-38, 42-50
10	41-50	1-10	11-21, 42-50		
11			11-21	1-10, 20-29	
12	1-8, 20-30			1-10, 20-29	
13		11-30	42-50		1-10
14	20-30		33-41		1-10

APPENDIX 4: EQUATING DATA SECOND STUDY, PART TWO

A	B	C	D	E	F	G	H	I
LANG	EXYR	% RIGHT	% RIGHT	D - C:	MEAN	CUT-OFF	CUT-OFF	H - G
		EXPOP	REFPOP	MEAN	PREVEX	EXYEAR	EQUIV	EQUIV
				REFPOP	-REFEX		TO REFEX	
			-EXPOP					
GERD	1982	64	61	-3	-9	55	53	-2
GERD	1984	75	67	-8	-3	59	59	0
GERD	1986	69	67	-2	-3	59	59	0
GERD	1988	66	63	-3	-7	57	55	-2
GERD	1991	70	70	0	0	63	63	0
ENGC	1981	61	58	-3	-10	49	47	-2
ENGC	1983	68	66	-2	-2	59	57	-2
ENGC	1985	74	74	0	6	63	65	2
ENGC	1988	59	67	8	-1	49	59	10
ENGC	1991	68	68	0	0	59	59	0
ENGD	1981	50	54	4	-20	41	39	-2
ENGD	1983	69	64	-5	-10	59	51	-8
ENGD	1985	80	75	-5	1	63	65	2
ENGD	1988	70	67	-3	-7	59	53	-6
ENGD	1991	74	74	0	0	61	61	0
ENGH	1981	68	68	0	5	59	61	2
ENGH	1983	67	65	-2	2	57	57	0
ENGH	1986	67	66	-1	3	57	59	2
ENGH	1989	65	61	-4	-2	53	53	0
ENGH	1990	64	63	-1	0	55	55	0
ENGV	1984	80	75	-5	1	63	65	2
ENGV	1985	71	68	-3	-6	57	55	-2
ENGV	1986	68	70	2	-4	57	57	0
ENGV	1988	70	69	-1	-5	59	55	-4
ENGV	1991	74	74	0	0	63	63	0
FREV	1981	69	63	-6	-6	59	51	-8
FREV	1984	77	72	-5	3	63	61	-2
FREV	1986	69	68	-1	-1	57	57	0
FREV	1989	72	71	-1	2	61	59	-2
FREV	1991	69	69	0	0	57	57	0
DUTC	1986	60	53	-7	-13	50	41	-9
DUTC	1987	64	53	-11	-13	51	41	-10
DUTC	1988	52	45	-7	-21	39	33	-6
DUTC	1990	67	62	-5	-4	49	44	-5
DUTC	1991	67	66	-1	0	49	49	0