GEOGRAPHY ENTRANCE EXAMS IN THE CZECH REPUBLIC IN 1997

Dana Řezničková

Department of Social Geography and Regional Development, Faculty of Science, Charles University, Prague, Czech Republic

Abstract: This paper covers research the aim of which was to evaluate the standard of geography entrance exams in the Czech Republic and, on the basis of established facts, judge the significance of the exams in terms of selecting applicants for university geography study.
The subjects of the study were eight entrance exams used during the selection process in 1997. Natural science students at Charles University answered the test questions as the first and then five university educators also answered them. These professionals at the same time evaluated individual tests according to previously established methodical instructions. The results of the students' performance and the statements of the panel on individual tests were the bases for relatively extensive analysis. The professionals' opinions on the content aspect of the tests have considerable value as testimony.

Key words: geography entrance exams, statistical analysis, qualitative analysis

1. THE AIMS OF THE RESEARCH

The results of written geography entrance exams with the results of a following background of the subjects present the determining criterion of the selection of applicants to pedagogic and non-pedagogic geography study at all nine universities in the Czech Republic. An urgent need has arisen to inquire into not only determining the characteristics (knowledge and abilities) of applicants for whom the geography test is designed, but also the quality and property of the test itself, resp. its conception.

The subjects of research were eight entrance exams used during the selection process in 1997 (the ninth test was sent late and, therefore, not included in the overall statistical analysis). Our purpose was to evaluate with qualitative and quantitative

methods the standard of individual entrance tests, and on the basis of established facts, judge the tests' significance in terms of the process of selecting applicants for university geography study.

2. WORK METHOD

The work method selected corresponds to the determined aims. In the course of three weeks eight entrance exams tests were given to thirty to thirty-five member groups of first year students of pedagogy at Charles University School of Natural Science and, at the same time, evaluated by five professionals according to previously determined criteria.

The given tests, the panel statements and all calculations for individual tests are given in the Final Report of Grants FRVS titled "Graduation Exams and Entrance Processes for Natural Sciences", Charles University, Natural Sciences, Charles University, KSGRR Prague, 1998. As the volume of contributions was limited, here are submitted only the basic information.

2.1. Qualitative evaluation of entrance exams

Five university educators with various geographic specialisations, some of them also having secondary school practice, delivered qualitative analysis of tests. The professionals evaluated in writing part of the essays and the general quality of eight entrance exams according to consistent methodical instructions. Specifically, they were asked to judge:
* the technical quality of the questions
* the correct solution of the questions
* the suitability of content to the specific knowledge of the discipline
* the questions representation of the content
* the standard of the required thought
* the aim of the questions or whether it tests knowledge or the student's teachability
* the difficulty of the tasks

The professionals judged the overall standard of the tests from these angles:
* the proportion between thematic wholes
* the representation of questions whose solutions required higher thought
* the content validity of the tests. This involves to what extent the test, as a whole, fulfils the aim.

2.2. Quantitative analysis of the tests

The results of the students' performance and the panel's statements for individual tests formed the bases for a relatively extensive statistical analysis. For each entrance test an *item analysis* and a ***reliability analysis*** were separately performed. Among others, an ***item analysis*** was performed, which traces the ***general exam data*** (the average number of points earned in the written test, the range of test results...), which would be attained ***with the omission of items***. In general, it becomes apparent that the most important data of these parts and the item analysis is ***the correlation between the item and the test***, which tells us to what extent the respective item contributes from the angle of what the given test measures.

On the basis of statistical method, we gained in individual test items with the best qualities. These, represent the test's conception, however they state nothing about the correctness of the general conception of the test. No wonder the professionals' statements, pursuant to the content standard of the test, cast some doubt on the value of the best obtained questions (e.g. a given question was ambiguous, a question demanded too detailed topographical knowledge etc.).

2.2.1. Reliability analysis

Let us direct our attention to the coefficient of reliability, which was calculated with the help of the so-called Cronbach α - coefficient and the split-half method. For both data it holds that the higher they are, the more reliable the test. The coefficient (see the graph) tells us if it succeeded in collecting items, which may be said to test to a certain extent the matter (knowledge and skills) and that the matter is tested reliably (i.e. with little error.) The high reliability of a test is certainly a property of its internal construction and still tells us about whether the test measures what it should. Partially reduced reliability can be then caused by different types of themes, which minimally overlap, and are included in the test. Therefore, it is important in comparing the reliability of tests to take into consideration not only the internal division by subject but also by type (representation of knowledge, application and interpretation of tasks), which were carried out by the professionals. On the basis of their judgement, thanks to the substantial variability of the content and type of questions, test #8 had the least reliability. On the contrary a higher level of reliability an 80% and above representation of questions requiring only a knowledge of topographical names and geographical knowledge in tests # 4, 6, and 7. This consistency of questions increases the reliability of a test (shows that the occasional mistake was incidental), not always its reliability. For example, a reliable test of the topographical names of a certain area is not a valid test of the understanding of the main characteristics of that area, much less a test of teachability. Another source of mistakes can be the ambiguous formulation of questions and their evaluation. As the following table shows, the first mentioned factor most likely influenced the level of reliability in tests # 3, 5 and even 7. As a whole, except for two tests, the share of ambiguous questions is a hindrance, because this fault on the part of the tests' creators can result in faulty answers from those tested.

Table 1 Characteristics influencing the reliability of the tests

Test #	Total amt. f items	Share (%) of ambiguous items in test	Maximum amt. of points	Difference in importance of items (min./max. amt. of points)
1	21	33	400	5 / 50
2	14	36	60	2 / 10
3	10	40	36	3 / 5
4	44	38	75	1 / 4
5	25	40	25	1 / 1
6	30	13	30	1 / 1
7	40	40	84	1 / 6
8	30	6	85	1 / 6

The higher reliability especially of tests # 4 and 7 could even have been influenced by the higher number of puzzling questions in the tests. Thus, the likelihood rises that the performance of those tested is not incidental.

The sore spot of the geography entrance exams except for tests # 5 and 6 is revealed to be the method of evaluating the tests, resp. assigning points to some items. Not only did the method of evaluating items influence the reliability of the tests, but it influenced the overall quality of the tests in a direct and substantial way.

In the evaluated tests, it is possible to discover different methods of assigning an item's point value. An item was worth more when

• its solution demanded more time-this was often the problem of open questions

• it included more choices (sub-questions)

• it contained more true choices

• the test's author considered an item more important, difficult in terms of content

It can be said that in the first three methods, the form of the question is decisive, in the last method the question's content, resp. its importance depends on the "opinion, feeling" of the test's author. As the sample below shows, this method of evaluation is considerably controversial and thus negatively influences the objectivity of the test.

For example, in question 3 from test 1 the answer is worth maximal 10 points. The question was: The pilgrim city of Muslims is

a) *Riyadh*

b) *Mecca*

c) *Jerusalem*

d) *Istanbul*

In the same test, a correct answer to question # 2 was assigned a maximum point value of only 5. The question was: In tropical climate these soils in particular are found

a) *black-earth and brown-earth*

b) *red-earth, yellow-earth and laterites*

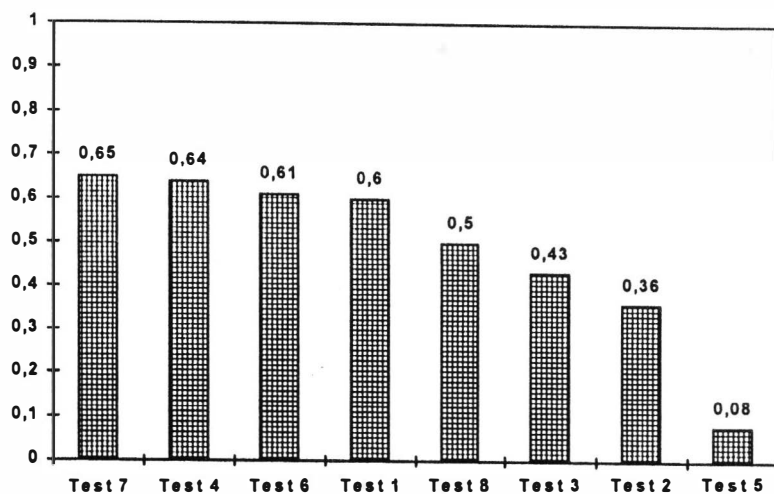c) *sandy clay and grey-earth*
d) *terra rosa and salty soil*
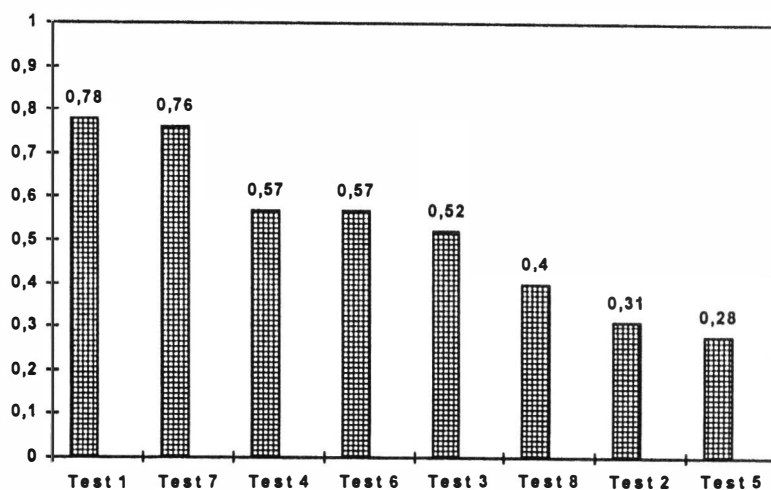

Fig. 1 Cronbach alpha

Fig. 2 Split-half reliability

2.2.2. Content analysis of tests' criteria

In this case, a content item analysis of all eight tests was again carried out but this time the items are not applied to the results of the relevant test, but to the external criterion representing the required geographic knowledge and skills. This criterion designated as the **sum** originates in the sum of the results gained from individual tests. So that multiple item tests were not over valued in such a summary, the overall result of a test was divided by the maximum possible points in the test-thus in an individual test each test subject could earn a maximum of 1 (all answers being correct) and a minimum of 0 (all answers being incorrect.) The results of this item analysis serve to correct the preceding item analysis. Items with a negative correlation and with correlation approaching zero are revealed as unsuitable predictors of geographical knowledge and skills on the premise that the "sum" reflects the representative criterion (quality) of the required geographic knowledge and skills. In light of the fact that the correlation are obtained from the results of only fifteen test subjects who filled out the eight tests, it is feasible as a statistically and unambiguously **conclusive predictor** of geographical knowledge and skills ($p=0.05$ and below) to take items in which the correlation **exceeds a value of 0.5**. In the following Table, it is shown that test # 7 contains relatively the most frequent occurrence of defined predictive items and, on the contrary, tests # 2 and 5 have none at all. These tests, on the other hand, consist largely (89 and 69% resp.) of unsuitable predictors of geographic knowledge and skills.

Table 2 Representation of questions with conclusive predictors of geographic knowledge and skills

Test #	Total amount of items	Share of items (%) with correlation		
		0,5 and above	0,2-0,499	Negative and 0,0-0,199
1	21	10	33	57
2	14	0	11	89
3	10	20	50	30
4	44	11	46	43
5	25	0	31	69
6	30	10	40	50
7	40	40	50	38
8	30	10	40	50

Of the total figure of 214 items only 9% (20 questions) reached a correlation of 0.5 or more, or it may indicate as *statistically* conclusive predictors of geographic knowledge and skills, resp. the contributed items in light of the fact that they measure the designated criterion. Their content aim is shown in the following summary.

From the content analysis of the twenty stated items among others it arises that we are dealing with questions designated by the professionals as of average importance, requiring a command of specific knowledge and even some higher thought. As regards teachability, most of the items test only partially. It is necessary to call attention to the fact that some of them are set ambiguously. Still, these items are revealed as the most

Table 3 Items with conclusive predictors of geographic knowledge and skills

Test #	Question #	Correlation	Value	Premises	Level of thought	Subject of question
1	9	0,509	2	2,8	M	Defining the state with the most forest.
	19	0,796	1,8	2	Z+M	Matching topographical names of cities, rivers, mountain ranges to specific areas.
2	x	x	x	x	x	
3	4	0,529	1,4	4	Z	Defining the census frequency,
	9	0,716	1,4	3,4	Z	Matching topographical names to certain objects (mountain, city ,,,)
4	4	0,739	1,4	3	Z	Stating the world's main wool producers,
	7	0,519	1,6	3,2	Z	Naming the 4 islands of the Great Sund and their state,
	33	0,578	2	2,8	Z+M	Finding Caracas on a map,
	36	0,529	2	2,8	Z+M	Finding the highest mountain in South America on a map,
	42	0,577	2	2,8	Z+M	Finding the lowest point of the Czech Republic on a map,
5	x	x	x	x	x	
6	6	0,579	2,2	2,6	Z+M	Defining a climate zone,
	11	0,518	1,2	3	Z	I ndicating the falsehood-location of Celtic nation, denominations in Croatia, illiteracy in Nepal,
	14	0,509	1,6	3,2	Z	Indicating the largest exporters of fish,
7	7	0,509	1,8	3	Z	Defining the state with the most rainfall,
	12	0,612	1,4	3,2	Z	Stating the object and its location according to topographical names,
	18	0,506	1,8	2,6	Z	Ranking states by population,
	27	0,627	1,8	3	Z	Defining in which states lie Alaska, Saar, Carinthia, Halich,
	39	0,556	2,7	2	Z	Choosing local cities,
8	15	0,526	2,4	2	M	Indicating the falsehood-the Nile's highest water level, the largest river confluence of the Wisla, The Volga the largest river in Europe, the elevation of the snow line above sea level, the elevation of Baikal below sea level, the total average rainfall on Milesovka v, Praded,
	16	0,620	1,4	2,4	M	Defining the cities through which the Volga does not flow,
	18	0,607	1,4	2,4	M	Defining the cities lying on the first Czech rail corridor,

productive in terms of what the criterion measures. We believe that on the basis of the content analysis of the twenty most productive items, we must cast some doubt on even the value of the external criteria.

2.2.4. Validity of the tests

The main property of the test or its validity (currency, adequacy of the test) was estimated for criteria and the contents validated. The first (i.e. criteria) estimation of validity is de facto item analysis, where there are individual items of pertinent tests which are applied to the results of a topographical super test which would originate with the fusion of entrance exams.

From the collected data it is perceptible that a test arising from the fusion of all entrance exams is reliable (*Split-half reliability: 0.833 and Cronbach alpha: 0.713.*) In this light, we may consider them suitable external criteria for the validation of tests (items.) The closest correlative affinity between test and criteria is shown by tests # 7 and 4. However, the unsolved question remains, whether the determined criteria have a predictive function in relation to university geography study.

Conceptual and construction validation consisted in the fact that five geography experts evaluated items step by step in all tests on a scale of five degrees in light of whether they are valid or not. Considering that its possible to judge the validity of items either from the standpoint of estimating the level of student knowledge or from the standpoint of estimating the teachability of the student in geography, the above mentioned expert judgement was applied twice- first the results were denoted by concept of *knowledge*, and then the *teachability*. On the basis of expert appraisal there was calculated for each test Aiken's coefficient of validity concerning rather the conceptual and structural validity of the test.

In the event that the purpose of geography entrance exams was to ascertain the knowledge which was vital to success in university study, then entrance exams are relatively valid (see the following graphs.) To a certain extent however, the result is cast in some doubt by the ambiguously set question in the methodical instructions.

The experts were provoked by question g into judging to what extent individual questions test geographic knowledge (quite certainly, with minor reservations, except in part, they rather do not test this-not at all.) This question does not unambiguously way what is understood under the concept 'geographic knowledge', resp. what we appreciate. The question is about the knowledge presented with geographic instruction in primary and secondary school in contrast to the knowledge required in university geography study or do we judge knowledge in terms of the level of inter-subject affinity?

As the following illustrates, according to the experts a relatively high validity in terms of teachability (the dark column) was reached by test # 8 followed by test # 1 and on the other end, the lowest, test # 3.
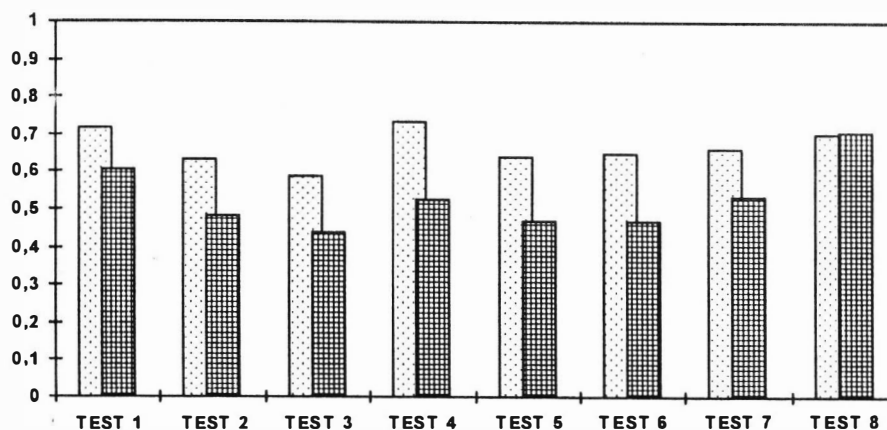
Fig. 3 Aiken's Coefficient of Validity

3. CONCLUSION

On the basis of the implemented statistical analysis and expertly judged entrance exams, it is clear that:

• by means of most of the evaluated geography entrance exams the knowledge of applicants is ascertained and subsequently evaluated. In many cases, of course, their selection, resp. the representativeness of questions is moot. Some questions the university experts, admittedly, were unable to answer correctly, they were uncertain and so had to look the answer up in literature. It seems that not all demanded knowledge is vital to success in university geography study. In any case, it would be misleading to be inspired with a whole set of items of evaluated entrance exams.

• In the framework of the applied analyses, a certain methodological inaccuracy was reached, which could have a distorting influence on the final result (e.g. students already accepted to university were tested, not applicants; criteria content analysis of tests was calculated only on the basis of 15 test subjects; experts, while evaluating individual items and overall standards of tests did not follow all determined instructions, often they had conflicting opinions, some questions in the methodical instructions were not defined unambiguously etc.).

• Even if all the entrance exams were probably conceived with the same general aim and preceding from the same pedagogic documents (educational standards, curricula), they still had more or less different concepts, content and structure of items, likewise a different level of criteria, concept and content validity. The order of 'the most' tests' changes depending on the method used. On the basis of the results of mathematic-statistical analyses or the method of abstracting, the content tests # 1, 4, 6,

295

mathematic-statistical analyses or the method of abstracting, the content tests # 1, 4, 6, 7 are revealed as relatively unsuitably formulated tests and the most problematic tests # 5 and 2. On the basis of expert judgement the most suitably conceived tests are # 8 followed by # 1.

◆ The evaluation of the quality of tests to a certain extent depends on what we consider 'quality' and how we look at it. In other words, first it is vital to decide what we measure, then how to measure it and, finally, interpret the attained results. Therefore, even in these studies commentary on the results of the mathematical-statistical method is relatively restrained. They will have more weight when we have clarified a profile of the applicants (what we will value more-intellectual abilities, exceptional knowledge, or the expression of the applicants' abilities etc.).

◆ Inconsistencies between the results of the quantitative and qualitative analyses draws attention not only to the danger of the use of statistical method as an end in itself, but also to the unsolved basis of the conception of the question: What should be the primary aim of a geography entrance exam? Is it a test of results (achievement test) or a test of teachability (aptitude test?) What should we imagine under the concept of educational teachability? What qualities in an applicant can a geography test screen?

◆ Evaluating applicants on the basis of the results of entrance exams in a profile of subjects is narrowly aimed at evaluating a student's academic capability at the moment (knowledge and intellectual skills.) Although it is generally true that to be successful student needs more than academic knowledge alone. The result of a test of 'academic geographical knowledge' and the real abilities of an applicant need not correspond at all. A student should understand knowledge correctly and use it appropriately, but also be creative, communicative, efficient, emotionally stable etc. How do we determine these characteristics? Social factors and emotional intelligence are the most difficult to map, and, therefore, are not taken into consideration during the selection process. Geography tests, often empowered by their lack of concept, have, of course, a limited capacity as evidence. It is necessary to consider other criteria in selection so that we may at least weaken the influence of the results of unsuitably conceived tests.

References

Bičík, I. a kol. (1998): Maturitní zkoušky a přijímací řízení z přírodovědných předmětů. Závěrečná zpráva grantu FRVŠ. KSGRR Př UK Praha.

Byčkovský, P. (1984): Základy měření výsledků výuky. Tvorba didaktického testu. České vysoké učení technické, Praha, 149 s.

Hniličková, J., Josífko, M., Tuček, A. (1972): Didaktické testy a jejich statistické zpracování. Státní pedagogické nakladatelství, Praha, 200 s.

Hrabal, V., Lustigová, Z., Valentová, L. (1994): Testy a testování ve škole, Pedagogická fakulta UK, Praha, 101 s.

Komenda, S., Mazuchová, J. (1995): Tvorba a testování testu. Lékařská fakulta Univerzity Palackého, Olomouc, 110 s.

Půlpán, Z. (1991): Základy sestavování a klasického vyhodnocování didaktických testů. Nakladatelství Kotva, Hradec Králové, 148 s.

Říčan, P. (1977): Úvod do psychometrie. Psychodiagnostické a didaktické testy, Bratislava, 192 s.

Resume

Přijímací testy na geografické obory v České republice v roce 1997

Příspěvek informuje o výzkumu, jehož cílem bylo zhodnotit úroveň přijímacích testů z geografie a na základě zjištěných skutečností posoudit význam testů v rámci výběrového řízení uchazečů o vysokoškolské studium geografických oborů v České republice.

Byla provedena kvalitativní a kvantitativní analýza osmi přijímacích testů použitých při výběrovém řízení v roce 1997. Na otázky v těchto testech odpovídali studenti 1. ročníku PřF UK a pět vysokoškolských pedagogů. Tito odborníci zároveň posuzovali jednotlivé testy podle předem stanovených metodických pokynů. Výsledky výkonů studentů i výpovědi posuzovatelů za jednotlivé testy se staly základem pro rozsáhlou statistickou analýzu. Mimo jiné byla vypočítána položková analýza a analýza reliability za jednotlivé testy i souhrnná položková analýza. Kriteriální položková analýza, kdy všechny testové položky byly vztahovány ke kritériu, resp. k součtu výsledků dosažených v jednotlivých testech, posloužila ke kriteriální validaci testů. Na podkladě expertních odhadů byl také pro každý test vypočten Aikenův koeficient validity postihující pojmovou a konstrukční validitu testů. Značnou vypovídací hodnotu má samozřejmě i srovnání názorů odborníků na obsahovou stránku testů.

Zjištěné výsledky této poměrně rozsáhlé analýzy především dokládají množství nevyřešených koncepčních a metodologických otázek souvisejících s funkcemi a tvorbou přijímacích testů ze zeměpisu.