

Georg Weissenro

Test tasks for Civic Education – Results of a pilot study

(in: Society for Civic Education Didactics and Civic Youth and Adult Education (publ.) (2005): Test tasks and evaluations on Civic Education, Schwalbach, Germany, p. 41 - 60)

1. What is new in the competency tests?

The educational standards of the Society for Civic Education Didactics and Civic Youth and Adult Education (GPJE 2004) are conceived as output standards and could therefore be used as an educational policy control tool through standardised tests (Abs 2004). However, it is the task of the KMK (The Standing Conference of Ministers of Education and Cultural Affairs) to decide on employing a commission for to work out national educational standards. The draft of the GPJE provides stimuli for this group that is still to be employed as well as for groups which are developed for core curricula at state-level. Since the draft on educational standards is borne by the policy didactics, it provides a weighted requirements catalogue with which the discipline can continue working in a sensible manner up until a decision is made by the KMK. In this sense the following report attempts to implement and develop further the requirements formulated by the GPJE. The process of developing domain-specific educational standards is at its beginning just as the development of suitable learning activities and test tasks. A new thinking on the quality of the teaching of political science has commenced with the reorientation that is intended by the educational policy from the input to output control of the educational system. The effects of the teaching on the pupils are once again the centre of attraction, for ultimately the school is there for the pupils.

There were already in the 1970s numerous attempts made to document systematically the output of the educational system. At that time one was involved with performance measurement within the course of the curriculum reform.

“Educational scientists have turned in recent years more and more often to questions of teacher rating so that the problem of teacher ratings has become a specialised area within the educational diagnostics, which in a special way has been penetrated by the personality and social psychology. The research into this area today uses complicated statistical procedures, which means that there exists a barrier to understanding in respect of the communication between the researchers and educationalists in the everyday school practice.” (Schwarzer/Schwarzer 1977, 9)

In 2004, the diagnostic teacher expertise continues to involve itself with these questions and is examining psychometric quality criteria for teaching personnel (Helmke 2003), but in addition to this, there is the question of implementing educational standards. Whereas it was still thought previously that the teacher kept himself a statistical record of his lesson in order to get an appropriate assessment of the pupil's performance, the focus of the current educational debate is more on whether the educational system can fulfil efficiently its work (Weißeno 2004b). The output shall not be alone measured by the teachers but in addition nationally and externally. However, the state of discussion of that time is in many points comparable with that of today, although there are different approaches. In the 1970s there were the “Soltau test tasks” of the working group on performance measurement in the *Volksschule* (elementary school) for the subjects of Mathematics and German, which the teacher could orientate towards and work with. Today, the task formats of PISA, TIMSS and IGLU provide the horizon to make comparisons. The wish for general reference dimensions is accordingly not new but is only now being implemented with blanket coverage by the KMK.

The first thoughts for performance measurement in civic education were expressed likewise in the 1970s. At that time the constraints on an empirically quantitative procedure had already been represented. “Future politically reasonable action of the citizen, targeted at civic education, lies outside learning success controls.” (Stenner 1979, 109). Still today the area of competency formulated in the educational standards of the GPJE cannot be tested as quantitatively valid. Moreover, this also applies for methodical capabilities because suitable procedures for sampling at national level have been lacking for this up to now. Only the ability to make political judgement can be tested nowadays as beforehand. In 1979 Stenner formulated the

following: “Insight embraces not only knowledge but requires as well forms of understanding and of analysis. In this sense the insights of the pupil can only be tested in application.” (113) This way can be pursued further working outward from the approach. It is true to say that the concept of knowledge is more precisely comprehended today and linked with capabilities. Weiner speaks of intelligent knowledge, which consists of knowledge, ability, understanding, action, motivation and so on. “Intelligent knowledge means a well-organised, disciplinary, interdisciplinary and practical-for-living network system of flexibly usable capabilities, skills, knowledge and metacognitive competencies” (Weinert 2000, 6). The difference between the modern concept of competencies and the concept of knowledge of Stenner becomes clear when the test tasks presented by him are examined. They are similar to tasks for learning objectives controls, such as a teacher could also give in classwork today, and which have their basis in a catalogue of knowledge.

An example:

“Which of the following measures do you consider as appropriate for convicted terrorists in a free democratic constitutional state like the Federal Republic of Germany:”

- surrender to the ordinary jurisdiction
- preventive detention on order of the police
- deportation abroad
- compulsory labour in penal institutions.“

(ibid.,127)

A single correct answer is here requested, subsequent to a series of lessons, as though as if it were an application of terms from the lesson. A standardised national test at the end of a grade (*Jahrgangsstufe*) tests in the meantime the ability to make political judgement without knowing about the teaching taking place in concrete terms in all the schools. But concrete contexts of application have to be pre-given here out of the politics in everyday practice. “Each illustration or putting into operation of a competency therefore must refer to concrete requirement situations.” (Klieme 2003, 60) This can take place in units which contain several questions on a (current) topic.

The test tasks, so-called items of a standardised test, let competencies be raised, in the learning objective controls solely a systematic structure of knowledge. This is an important difference which is mostly not observed in the educational debate because the items are presented as established control tasks out of the school. The competency test procedures document the domain-specific performance dispositions, but do not inquire about any catalogue of teaching materials and lesson contents. The solving of the task formats demand much more domain-specific action requirements, but which begin with the systematic structuring of knowledge. (Weißeno 2004a)

The task formats that are presented followingly take their orientation from the capability to form political judgement formulated in the educational standards of the GPJE (2004, 16). This capability can be shown in a variety of competency stages in applied situations. They tell something about which actions and mental operations can be carried out properly with a high probability. The task formats undertake “operationalisations” and provide in this way concrete information on the requirements to be expected (see Faulstich 2003, 99ff for basic criticism). In contrast to learning activities, they shall document the progressive complexity and difficulty, whereas the competency to make political judgement is subdivided into three degrees of difficulty (understanding, analysing, and judging). Peter Massing and Jessica Schattschneider make a difference between compatible areas in their contribution to this volume, which they subdivide once more additionally into stages in accordance with the number of points acquired. The interval scale undertaken here enables a competency spectrum to be demonstrated. The competency stages themselves have common statistical features: In PISA for example, it is assumed that the average learner will solve correctly within each competency stage an average task of this competency in 62% of the cases. The competency tests measure therefore differently than is done with marking via performance controls at school.

Through the codification of the answers, also the pupils who have given an almost correct answer obtain partial values of the maximum point figure. Thus the awarding of points is dependent on the respective level of understanding. Thereby the information contained in the answers are used more accurately and differentially than it is possible in most learning tasks. What is more, each item can be allocated to

several positions on the scale of difficulty. The competency model of the GPJE is a scientific construct. The implementation in the core curriculum, resp. teaching plan and in the lesson can in turn be geared to the test tasks. Insofar, the test tasks are specifications and enable the questioning in the lesson to orientate to the level of the competency stage. Test tasks and results are then relevant for the teachers in respect to being able to act and can contribute to the continual quality development through feedback of the level of competency attained.

2. The situations of use in the test tasks

Test tasks for civic education require reading literacy. Reading literacy is generally the ability “to understand written information for a large number of purposes, to use it and to reflect on it.” (OECD 2002, 28)

This also includes everyday information on politics. Being able to deal with information on political events at the level of everyday knowledge or with official documentation is part of reading literacy and is up to the 9th schoolyear already well developed. Therefore the reasons for not answering or a lack of concept in the answer can absolutely lie in the weakness in reading literacy. Detlef Oesterreich has referred to ruling out reading literacy with tests of knowledge and capability due to basic difficulty. If in a test knowledge it is decided on what is presented as a result of knowledge of the facts “it does not suffice if the correct conclusion can be likewise found with general intelligence.” (Oesterreich 2004, 252) Furthermore the relationship of indicator and what is indexed in the case of a test of capability in the area of civic education is not simple to determine. Therefore it appears doubtful to those responsible Germans in the Civic Education Study “whether most of the 13 items that are identified in the international study as skill items actually measure such a thing as political capability.” (ibid., 251)

Though no test of knowledge is intended with this pilot study, the problem has also still to be dealt with that is involved with the construction for the competency of being able to make political judgement. It was attempted to encounter this problem area from two different aspects.

- 1.) Since the reading situation, the structure of the text presented and the features of the questions that are set are important components of the reading procedure, domain-specific task formats are to be developed which do not represent any problem for the pupils in respect of their reading literacy. In order to achieve this, the units, which were revised over and over again, were tested one after the other in four classes beforehand. The pupils reported on their difficulties in the subsequent discussion in class and thus brought in their own interpretation into the revision.

- 2.) The assumption is of domain-specific importance that only knowledge ensures the understanding of the applied context that is being presented. It shall not be possible to solve the tasks with everyday political knowledge. Subject terminology, which is not a matter of course in the everyday language, is therefore taken into account. The tasks have accordingly to be formulated in such a way that they are not only to be solved by thorough reading of the text but surpass just the understanding of the text. Therefore tasks are worked into the units that test whether the pupils have perceived the specialist associations. Capabilities as “recognising the significance of political decisions for one’s own life” or “reproducing in a structured way complex political facts” (GPJE 2004, 16) are hereby meant. The subject-specific text comprehension surpasses in principle a comprehensive general understanding of a text. To what extent this analytical differentiation has indeed good selectivity in the practice of structuring the tasks and with the answering of the questions, will have to be tested individually and discussed.

Test tasks are furthermore differentiated by the tasks formats on attitude measurement. Attitudes are often measured in electoral research, in research into political culture, in recent times as well in the context of civic education (Meierhenrich 2003). These task formats do not involve observed knowledge and capability but an ordered relation between for example “not at all relevant” over “not rather relevant”, “partially relevant, partially not”, “rather relevant” and “fully relevant”. The variables are here ordered in ascending order according to their valuation. Such ordinal-scale variables have an empirical significance in respect to their order. But shall the competency be determined, especially interval and ratio levels are helpful. In this

case the difference between two values is accorded an empirical significance. The variables can also accept the value zero as the lowest conceivable value. The solutions range from totally wrong to fully correct, but also the mean values are sensible statistical core values for to describe the variables. The normal distribution can be furthermore decided on with these variables.

The task formats shall set the contents for an applied context. This is taken account of in the pilot study, whereas respectively current presentations of political affairs in the mass-media are chosen. From the level of the demand requirement, they have to be suitable for the target group of pupils of schoolyear 9. Therefore presentations selected from the *Bild-Zeitung* or the youth magazine of the Federal Agency for Civic Education – *fluter* – are principally suitable. The journalistic preparation of the issues shall ensure that a general text comprehension is possible at any time, but specialist understanding must be added. The knowledge and capability being used are more reduced by these media than would be the case with more demanding texts of newspapers such as *Frankfurter Allgemeine* or *Die Süddeutsche*. Moreover, texts on political events have been chosen which have been discussed publically over a lengthy period and which are still topical for pupils at the time of the questioning. The knowledge and capability acquired in the schoolwork are shown in the applied contexts which the pupils encounter every day.

It depends on whether the knowledge acquired in the learning situations remains passive or not. The imparting via artificially constructed hypothetical basic situations (Schelle 2003), via inventions and discoveries in human history within the context of a didactics of teaching arts (Grammes 2003) or via an institutional knowledge are not helpful, because the knowledge is reduced here to a catalogue of (basal) knowledge which is fully detached from the political reality or situation. The declarative knowledge (explicit, verbalised knowledge on issues) has to be increasingly proceduralised, that means being imparted in automatically accessible associations. Proceduralised knowledge must be reconstructed explicitly in applied cases, here out of the everyday political life. Cases from the everyday political life have never the same conditions and are always different.

In this connection reference has to be made to a further basic difficulty. If pools of tasks are to exist for every subject, new contexts must be continuously sought for civic education. Political events as applied situations quickly age, are always different and structure themselves differently outward from their base patterns. Civic education differs from other subjects, which don't know a problem like this. Physical phenomena of levitation and sinking can be observed in a lot of typical everyday situations, which are in addition not linked to any current everyday events. Historical applied contexts are likewise very much less dependent on the everyday politics. Insofar, basic patterns of the political affair as domain-specific scheme can be determined for the task formats in the subject "Civic Education", but they are again and again to be embedded anew in the current applied contexts. It is even more difficult when one wants to measure the competencies on an international comparison. Various countries have different styles of politics, traditions, patterns, and special features in democratic structures which are hardly comparable. The solution of the IEA studies to take hypothetical situations without any concrete applied relationship to the everyday political life is not a satisfactory solution from a domain-specific perspective.

Principally 4 types of *applied situations* can be differentiated:

1. politics for the private use (e.g. electoral voting, effects of political decisions on the personal planning for life),
2. the public (e.g. action as a citizen),
3. the professional (e.g. economic and political framework conditions) and
4. the educational use (e.g. for learning).

This can be supported in the lesson through case-based teaching-learning environments. Only when the teaching-learning material has already taken into account the diversity of the applied situations, the pupils can develop the different components of competency (Richter 2003, 121f). The handling of cases is checked in the test tasks and there are no isolated components of competency as e.g. factual knowledge (Klieme 2003, 66).

3. Selection and organisation of example tasks

The texts and tasks chosen for this pilot study reflect in extracts the spectrum of the degree of difficulty which the 4 types of political situations are to be assigned to. In order to simulate authentic political situations, the pupils shall a) ascertain political information, b) develop a politically-related interpretation, c) reflect on the political content and evaluate it (see OECD 2002, 33ff). All three aspects are important for the full understanding.

- *Ascertaining political information* requires the localisation of political contexts. The pupils have to find their way through ambiguities, designate the political key statement (understanding).
- Political understanding is required for the *developing of a politically-related interpretation*. The pupils have to process the political information that is given and differentiate it from other associations (analysing).
- The *reflection and evaluation of a political fact* requires the development of one's own standpoint from a subject perspective (judging).

The distribution of the 8 tasks in the pilot study looked like as follows:

Aspect	Tasks	Points/maximal	Points individual
Understanding	2	25	10 – 15
Analysing	4	80	20
Judging	2	80	40

There is only a choice of tasks presented in the present work. (Both units are published in their entirety on the internet on the pages of the *Institut für Testentwicklung und Evaluation der PH-Karlsruhe* (Institute for test development and evaluation of the university of education in Karlsruhe.) Moreover, both the questions on judgement are evaluated because this level of competency is of interest for the subject didactics. The question is asked whether the normative stipulations of the GPJE are too high here. At first glance many teachers consider these performance requirements as hardly attainable. The reason for this can lie in the fact that there is

a difference between learning and test tasks (Weißeno 2004c). The tests were conducted without the teacher. The interpretation of the data of the judgement questions was effected with a criterion catalogue. The points were awarded and controlled in two separate procedures. In the case of large deviations, the third evaluation was done by experts. Task formats like tables, diagrams, video recordings were not employed. Solely used during the pilot study were different sorts of text, an interview and an online report. Each task is presented according to the political situation, the competency stage and the text format. 4 tasks were allocated to each text. A text with tasks is described here as a unit. The texts and tasks are reproduced in the same form in which they were given to the pupils. The validity of content of the test is ensured through the three levels of competency which are measured within both the units. No reliability exists in this pilot study because measurement errors as clever guessing, attitude to test and so on are not to be ruled out and the number of tasks is too low.

Labour market and taxation reforms

This interview text (see below) concerns a newspaper text. It is classified as a text for public use from the political situation. The interview was conducted subsequent to an important vote in the *Bundestag* following months of public discussion. Both questions 24 and 25 refer to the labour market reform (Hartz IV), on which the government and the opposition have agreed in the parliamentary mediation committee. The interview was conducted at the start of February 2004 so that the events were still topical and the pupils could use their knowledge. The reasons for the inclusion of this text in the survey tool are: short, precise, comprehensibly formulated questions and answers; the situation in the *Bundestag* is an applied context in respect of the role of parliamentarians as part of the core concept of popular government and party democracy (GPJE 2004, 21).

Introduction:

The following questions relate to the compromises and voting on labour market and taxation reforms. The interview took place the day after the voting in German *Bundestag* . Read it through carefully and then answer the questions 24 to 27.

21. December 2003, Hamburg

Federal Chancellor Schröder in an interview with “*Bild am Sonntag*” of 21. December



BILD:

Mr. Federal Chancellor, 12 Red-Green representatives voted against your own labour market reform on Friday at the end of a long and hard year of reform. Is the Coalition capable of governing?

Schröder:

No doubt about this. When we needed the Chancellor majority – i.e. 302 votes –, it was there. When we needed the simple Coalition majority – i.e. we had to lay more votes on the scale than the total opposition together – we succeeded here as well.

BILD: How often may this voting conduct happen again?

Schröder: It is always easier if you have no dissidents in the Coalition.

BILD: The people are less relieved through the tax reform than it was planned. Much ado about nothing?

Schröder: I would have liked to have had the planned taxation of 22 billion Euros already by January 2004. This has failed due to the will of some CDU-ruled states – besides not due to Edmund Stoiber und Angela Merkel. But on the whole, we can live with the compromise.

Quelle: http://www.bundestkanzler.de/Interviews-.7716.580116/Federal_Chancellor_Schroeder_in_the_interview_with_Bild...htm

Question 24: Labour market and taxation reforms

A law is passed in the *Bundestag* when it is decided by majority vote.

Please cross your answer: During the voting on the labour market reform on 19. December 2003 the Chancellor majority was required because

- The Federal Chancellor demanded it from the parliament.
- All important laws are passed with the Chancellor majority.
- The compromise of the parliamentary mediation committee required this.
- CDU und CSU applied for it.

Frage 25: Labour market and taxation reforms

12 representatives of the government majority voted against the labour market reform. This is a problem for the government. The opposition parties say that the government has no longer its own majority and is no longer capable of governing.

A representative is selected as candidate for election by his party. He is only answerable to his conscience when voting.

Weigh up and justify how a SPD representative should behave relative to his own party when voting on the labour market reform.

The tasks set in the unit extend to all three competency stages. Question 24 relates to the aspect ascertain information of the competency stage: understanding. It requires applying in the context the knowledge on the differences between the simple and qualified majority (Chancellor majority). There are two features that make the solution of the task simple. Firstly the Chancellor majority is reached with 302 votes. Secondly these are more votes than the total opposition has. Who has ascertained this political information as being significant for the voting modality and knows from other cases that only in a few voting situations a qualified majority is required by the Basic Law, he cannot select solutions 1 and 4. To infer a 'demand of the Chancellor' from a general text understanding of 'Chancellor majority' is subject-wise just as misleading as the political misconception of giving more rights to the opposition parties than to the government parties in these matters. Solution 2 is very vague; still a correct paraphrasing which can at least indicate an understanding of the subject. Solution 3 requires knowledge of all the rules for the different applied cases.

Question 25 is the most difficult of the unit. For its solution the pupils have to refer to specialist knowledge on the selection of candidates, on the mandate of representatives, on parliamentary balance of power, on political interests and loyalties. It was necessary to weigh up between the question of individual conscience and the demands of one's own party to hold on to power, and then to formulate a political judgement on a current case. For this an estimation of the political significance of the voting was to be undertaken. The total number of points could only be attained with answers on the stated issues. The own opinion should not turn out to be striking in everyday language but should be formulated as a considered judgement on the use of specialist terminology. The basic conflict between a question of individual conscience and a whip (voting on party lines) had to be substantiated with consideration.

There were no answers that obtained the total number of points because the pupils included too few factual aspects in their considerations. This should still not be normally expected in the *Sekundarstufe I* (secondary school level I). There is the added fact here that two *Realschule* (secondary modern school) classes were interviewed. A further reason possibly lies in the fact that as far as we know the subject teacher in the lessons attaches large importance to the imparting of knowledge or factual knowledge. The influence of the person teaching remained unconsidered in this study.

More than 20 points were obtained as an example for the following answer:

“You should vote according to your own opinion but also take into account the opinions of others. If you vote against, you will have a bad conscience in respect of the Chancellor, although it is always better to be of one opinion in a party. If you vote for and not against, you will gain nothing and perhaps regret it.” (No. 31)

Answers were given zero points which represent an everyday expression of opinion:

“He must participate in the labour market reform and vote for it.” (Nr. 36)

“He should stand by his party and not be a dissident. He should represent the opinion of his party.” (No. 21)

Reform of the health system

The tasks of this unit are involved with private usage according to the situation. It is necessary to personally assess correctly the consequences of political decisions. The online newspaper text taken from a youth magazine deals with the effects of the planned (and then implemented) health system reform. The practice fees were especially being discussed publically at the time of the questioning, so that the political associations were present. The reasons for the inclusion of this text in the survey tool are: clear classification into various aspects; precise language and motivating layout; a transparent number of political players with different interests is indicative of political decisions being made; the effects of the health system reform and the behaviour of the citizens are brought into relationship; the text is ambiguous

in respect of the political conflicts and the economic preferences of the citizens and interest groups. The text creates a feeling of ambiguity by illustrating the economic cutbacks that are affecting everyone. The pupils have to first establish a connection between the cutbacks, which are possibly perceived as shocking, and the task of politics to provide structures. But there are a lot of linkage points in the text for this. The tasks set in the unit relate again to all three competency stages.

Introduction:

The following text deals with the effects of the reform of the health system which was decided upon at the end of 2003. It lists a series of economic consequences and refers to the discussions in the political arena. Read it through carefully, underline the important points and then answer the questions 1 to 4.

Becoming a luxury: a visit to the doctor?

Looking after our health is becoming expensive



Health is the most precious good. That's clear. Still when it comes to how much money we are prepared to spend on it, many people are beginning to waver. The government is also currently experiencing this when it attempts to convince the citizens about the reform of the health system.

At first glance everything seems to be good news. The new regulations lower the contributions to the statutory health insurance companies and thus the high incidental wage costs, which the economic location of Germany is reeling under. At the same time the burden on the health insurance companies is eased by around 10 billion Euros and they can then use 3 billion Euros directly to reduce their debts. But the problem is: the patient will have to pay more when he visits the doctor in the future.

Health insurance card and a tenner on the table

So what costs do I have to bear next year? The good news first: The general contribution rate – half paid by the employer, half by the employee –, of an average of 14.4% shall be lowered to 13.6% in 2005, and then to 12.15% in 2006.

The bad news: a visit to the doctor shall cost in the future 10 Euros per quarter* and per treatment. Who wants to save money, has to stay healthy. Consumer protection bodies are anything but satisfied with the changes being pursued. They criticise that the money which is saved from the employers and the insured who do not avail of any services has now to be paid by the sick. In a stroke the financial burden on the health system would not become lower – it is only being redistributed.

Glasses and teeth – stock up now ...

The health insurance companies are concerned about the forthcoming changes, too. The fear about the reform are driving up the costs even now. For the insured are stocking up with

glasses, dentures and medicine.

As a consequence of this hamster buying, the statutory health insurance companies mostly do not see themselves in the position to lower as required the contributions paid to them. Most of them want to wait and lower them in the course of the year; four of them even announced that they wanted to raise their contributions. The conflict with the government is thereby pre-programmed, for it is insisting that the health insurance companies are now legally obligated to lower their contributions.

Samira Lazarovic works as an editor on economics affairs for ntv., 1.12.2003

*Quarter: comprises three months

Source: Fluter, Bundeszentrale für politische Bildung (Federal Agency for Civic Education), <http://www.fluter.de>, 1.12.2003

Question 4: Reform of the health system

It can be discussed whether there is a fair social balance in the health system reform.

Do you agree with the position of some players that there is not a fair social balance in the health system reform? Weigh up the different positions in the text and justify your answer **in your own words**.

*player/s: persons and organisations involved

Question 4 falls into the competency stage: judging. To solve it, the pupils have to refer to specialist knowledge on how political players act and on the objective of social justice. It was necessary to find out about the players of the Federal Government, statutory health insurance companies/insured people, employers, consumer protection officers and relate in a considered way their positions on the question of social justice. The criteria for political actions were to be evaluated for this. The pupils had to draw conclusions from a subject-related perspective and justify them by weighing them up carefully. This had to be done without the text making explicit statements on them. A superficial argumentation could have resulted in a moral view on the politicians who demand only more money from the citizen. The

more profound consideration is founded on abstract ideas of politics (“reflecting on the consequences and side-effects of political decisions”, GPJE 2004, 16).

There were no answers that obtained maximum points because the pupils addressed too few subject-related aspects. This was not expected from schoolyear 9. But basically these requests appear reasonable because interconnections to the private sphere and to the national level must be drawn on in order to assess political decisions.

The following answer obtained more than 20 points:

“I agree with the players who say that there is not a fair social balance in the health system reform. Certainly the statutory health insurance companies are saving money, but the sick have to pay more for this. Therefore all would have the same outlay if the sick had a higher rate of contribution to pay as the healthy. And this is not right, for often the poor get sick faster because they simply do not have enough money to keep healthy. Following the new health system reform, they still have to raise this money even when they are sick.” (No. 23)

The ability to judge is clear in this answer. Pupils at this age level can judge; in doing so however they do not take into account all the information that was searched for. The answer works out the association of political framework conditions, the action of the players and the value of social justice.

Answers with everyday expressions of opinion were valued with zero points.

“Yes, I agree that that there is not a fair social balance in the health system reform.” (No. 6)

This is exactly the opposite of the answer required, although the expression of opinion matches the topic. But the opinion has neither been founded on a subject-related connection nor resulted from a process of careful consideration.

4. Individual results of the pilot study

Task 24 is the easiest in the questionnaire. 22.8% of the pupils gave a wrong answer. Solutions 1 and 4 are valued as subject-wise erroneous.

Labour market 24

		frequency	percent	valid percent	Cumulative percent
valid	0 points	13	21.7	22.8	22.8
	10 points	31	51.7	54.4	77.2
	15 points	13	21.7	22.8	100.0
	total	57	95.0	100.0	

Required for the correct answer are precise knowledge of the consequences and the understanding of the purpose of a parliamentary mediation procedure. Only 22.8% of the pupils have answered totally correctly. The result is interesting that 54.4% have crossed the solution 3 as approximately correct. This is therefore considerable because both the majority and the Chancellor majority are referred to in this task. Even when no explicit knowledge on the part of the majority of the pupils was present, it can still be presumed that implicit capability (Spitzer) has developed through the teaching. The pupils have used their knowledge on different majority requirements to find approximately the correct solutions in this applied situation. There is definitely present perceptive knowledge (Weinert) on the importance of different majorities and this has been applied. Since the values do not lie closely together and the task is solved totally correctly by only a few, it can be assumed that the task has not been too easy nor too difficult. The mean value is 8.86.

Proportion of males/females with the labour market 24 question

% of the sex

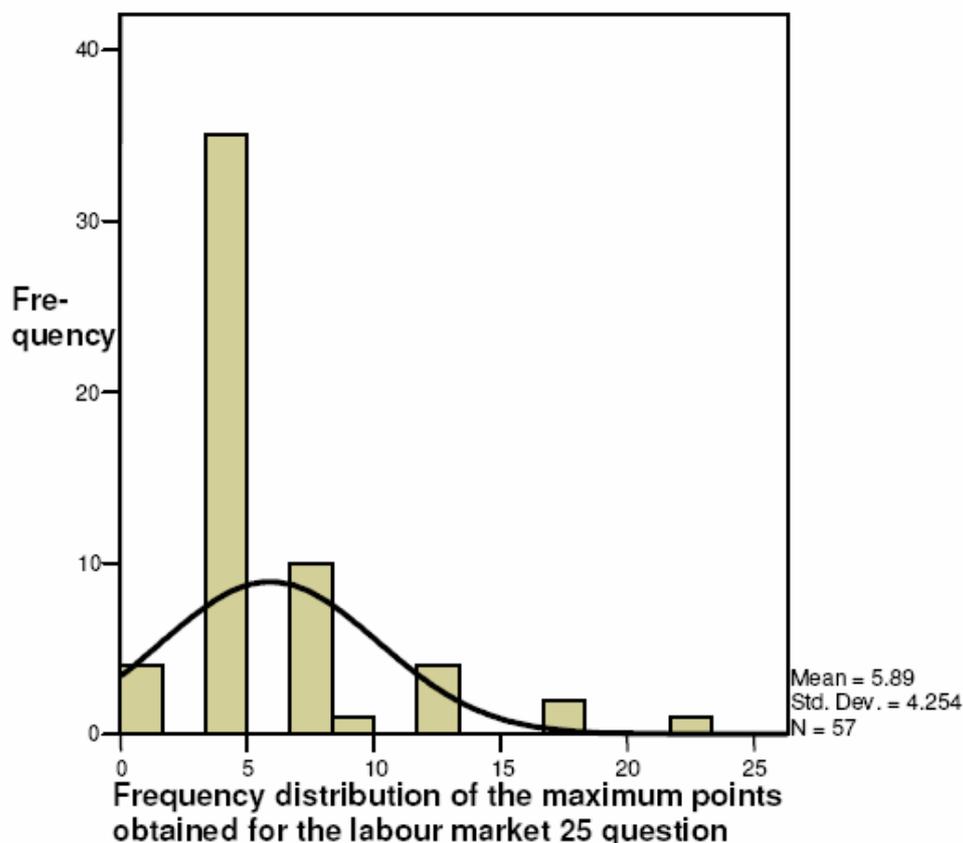
		labour market 24			
		0 points	10 points	15 points	Total
Sex	Male	25.9%	51.9%	22.2%	100%
	Female	20.0%	56.7%	23.3%	100%
Total		22.8%	54.4%	22.8%	100%

When examining the sex-specific results, then it has turned out that the girls have performed better than the boys in all answer dimensions. Theoretically seen, these results e.g. can thus be interpreted in such a way that the girls have learned more

intensively – a fact that a large number of studies from other disciplines refer to. The girls possibly are more endowed with faculties to react quicker to a new situation, However, the differences are not that large.

The results of the judging question on the voting behaviour (labour market 25) look like as follows:

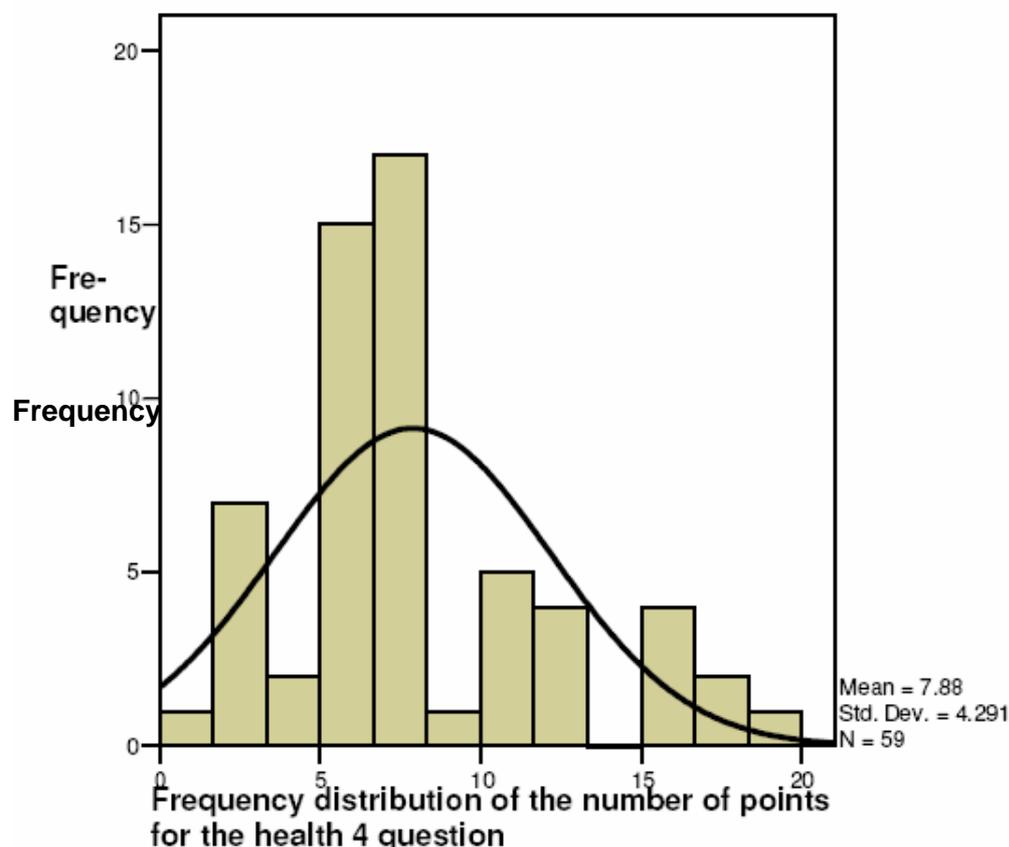
It is striking that nobody obtained the maximum number of points (40 points). One pupil obtained 22 points, 61.4% even less than 5 points. Owing to the large number of outliers, the distribution should, however, correspond rather only approximately to the normal distribution. The wide scattering of the raw values can reflect the actual differences of performance within the population, or still mean that the task is too easy or too difficult. Nevertheless, the boxplot (see below) has the tendency to show a similar picture.



Should it be attempted to interpret this result theoretically, several assumptions are possible which need to be examined in follow-up studies. It is conceivable that the pupils still do not have sufficient experience in making judgement. Even if – due to

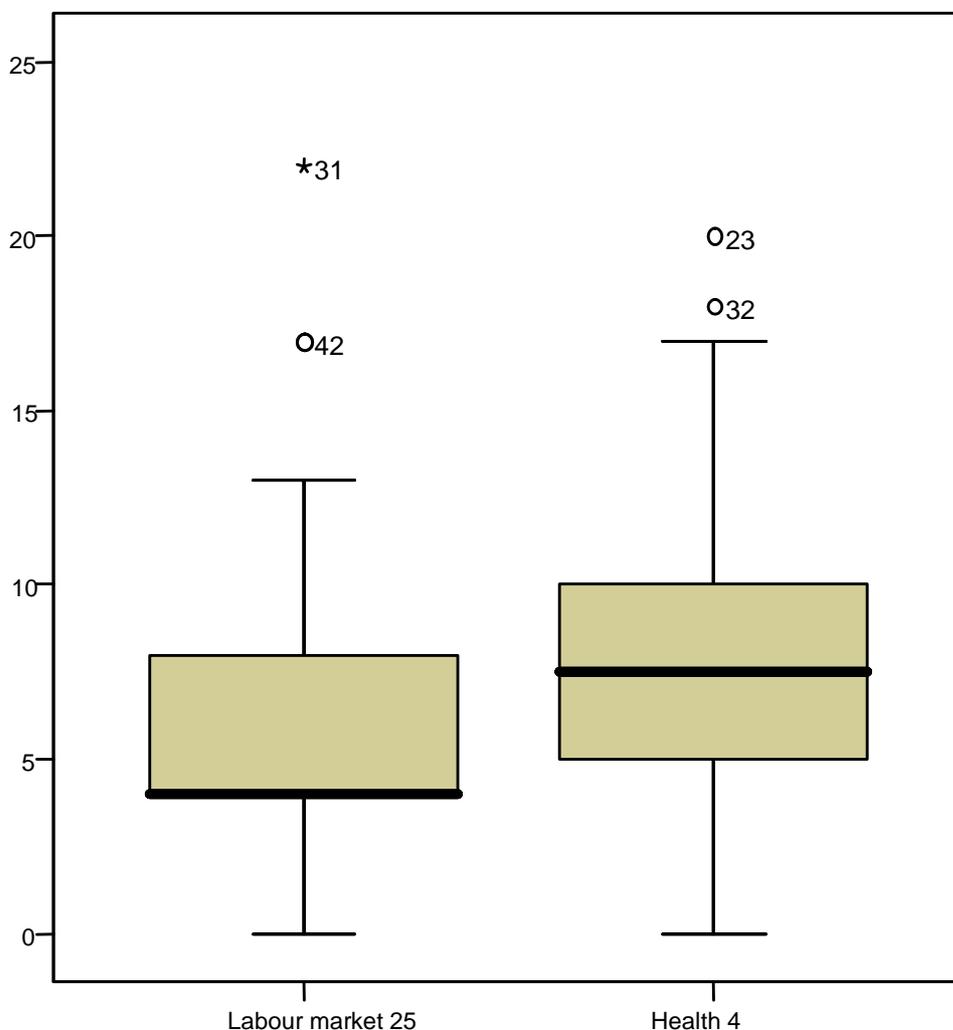
the maturing of the brain – a fully developed political value system on the basis of whose principles you had to communicate as regards content cannot be expected (Spitzer 2002, 358ff), the low number of points obtained by some pupils is still striking. If it is assumed that a considered political judgement is possible, as the answers with the higher number of points show, there is still presumably too little experience present. Teaching which demands only facts, or wants to build up a catalogue of various knowledge, does not enable any experience in judging to occur. Teaching in which the concise political opinion is asked for and is rewarded with good marks, likewise does not enable any experience in judging to occur. If the teachers also report back that they consider the task formats for judging too difficult, this estimation perhaps is furthermore to be explained by the fact that the development of the brain – especially the cortex – has not been completed. It may therefore be difficult to foster political judgements in the teaching. We interpret the answers with the higher number of points in such a careful way that political judgements are basically possible. We did not expect a completely correct answer.

The results of the judging question on the health system reform look like as follows:



The results are similar here, but a little bit better. Anyhow 21.7% of the pupils obtained up to 5 points and 76.3% up to 10 points. The maximum number of points is 20. Considering the data on the basis of both school classes, then the result of the Kolmogorov-Smirnov test is ($9a=0.192$; $9b=0.202$) that the critical values of 0.246 and 0.242 are not exceeded. The values given can therefore be regarded as normally distributed.

The results of both variables are visible through the boxplots. The scattering is in the case of the health variables lower in their extreme values than with labour market variables; therefore it is meaningful.



Boxplot for the labour market 25 and health 4 questions

On the whole it can be stated that the pupils have difficulties in judging. In order to narrow down further the results on the basis of the data given, we have collated together the percentage values in respect of the questions on each competency stage (understanding, analysing, judging) and compared them with each other.

Mean values of correct answers in percentage of the three levels: understanding, analysing, judging

	Mean value	N	Standard deviation	Standard error of the mean value
Understanding	61.6667	60	30.20645	3.89964
Analysing	57.3958	60	20.30917	2.62190
Judging	16.6875	60	8.81130	1.13753

A comparison of the percentual correction values of the questions on understanding and on analysing is not significant because the mean values are anyway approximatedly the same. But the other two comparisons are significant. The non-parametric test of Wilcoxon that is undertaken to compare the data comes to the same results. Accordingly, neither understanding nor analysing are dependent on judging. The pupils answer questions on understanding and analysing significantly better than questions requiring judgement. Neither understanding nor analysing is a sufficient condition for judging. The data therefore discloses nothing about what judging is dependent on. Furthermore, it remains an open question what has to be present to ensure that judging can occur. Follow-up studies would have to clarify whether this is attributable to the motivation of the interviewees, the lack of practice or other reasons.

In terms of subject-related didactics, the question is still interesting whether the difficulties with judging are caused by the teaching offering too few applied contexts for exercising judgement. Individual studies from the qualitative research into teaching suggest this conclusion (Kuhn 2003; Massing/Weißeno 1997). The fact that the pupils have clearly better solved the questions on understanding and analysing could indicate that they have little experience with this requirement. On the other hand, the results can be age-dependent, but this cannot be clarified by this pilot study. No comparative data at all could be drawn on for this pilot study. Furthermore, it is not to be excluded that the task format was too difficult. However, there exist on

the other hand no indicators that the norm for being able to make political judgement should not be fulfilled by the draft paper of the GPJE on educational standards. Insofar, there are at the moment more questions than statistically-substantiated answers. More research continues to be needed here.

Bibliography

- Abs, Hermann Josef 2004: Weichenstellungen in der Einführung von Standards im Fach politische Bildung, in: Politische Bildung, Heft 3, S. 82-94
Abs, Hermann Josef 2004: Indicators for introducing standards in Civic Education, in: Civic Education, Issue 3, p. 82-94
- Faulstich, Peter (2003): Bildungsstandards, Kompetenz und Bildung, in: GPJE (Hrsg.): Politische Bildung zwischen individualisiertem Lernen und Bildungsstandards, Schwalbach, S. 94 – 107
Faulstich, Peter (2003): Educational Standards, Competency and Education, in: GPJE (publ.): Civic Education between individual learning and educational standards, p. 94 – 107
- Gesellschaft für Politikdidaktik und politische Jugend- und Erwachsenenbildung (GPJE) (2004): Nationale Bildungsstandards für den Fachunterricht in der Politischen Bildung an Schulen. Ein Entwurf, Schwalbach, Ts.
Society for Civic Education Didactics and Civic Youth and Adult Education (2004): National Educational Standards for teaching civics in schools. A draft paper, Schwalbach, Ts.
- Grammes, Tilman (2003): Handlungsorientierung im Politikunterricht, in: Politik unterrichten, Heft 2, S. 13 - 18
Grammes, Tilman (2003): An orientation for action in the teaching of politics, in: Teaching Politics, Issue 2, p. 13 - 18
- Helmke, Andreas (2003): Unterrichtsqualität, erfassen, bewerten, verbessern, Seelze
Helmke, Andreas (2003): Documenting, assessing, improving the quality of teaching, Seelze
- Klieme, Eckhard u.a. (2003): Zur Entwicklung nationaler Bildungsstandards. Eine Expertise, Frankfurt
Klieme, Eckhard i.a. (2003): The development of National Educational Standards. An expertise, Frankfurt
- Kuhn, Hans – Werner (Hrsg.) (2003): Urteilsbildung im Politikunterricht, Schwalbach
Kuhn, Hans – Werner (publ.) (2003): Forming judgement in the teaching of politics, Schwalbach
- Massing, Peter/Weißenö, Georg (Hrsg.) (1997): Politische Urteilsbildung. Zentrale Aufgabe für den Politikunterricht, Bonn/Schwalbach
Massing, Peter/Weißenö, Georg (publ.) (1997): Forming political judgement. The key role in the teaching of politics, Bonn/Schwalbach
- Massing, Peter/Schattschneider, Jessica Simone (2005): Aufgaben zu den Standards der Politischen Bildung. Ergebnisse einer Pilotstudie, in: GPJE (Hrsg.): Testaufgaben und Evaluationen in der politischen Bildung, Schwalbach (in diesem Band)
Massing, Peter/Schattschneider, Jessica Simone (2005): Assignments for the standards of Civic Education. Results of a pilot study, in: GPJE (publ.): Test tasks and evaluations on Civic Education, Schwalbach (in this volume)
- Meierhenrich, Volker (2003): Wie können Schüler politisch urteilen? Kategorien politischer Urteilsbildung im Bewusstsein von Schülerinnen und Schülern, Schwalbach/Ts.
Meierhenrich, Volker (2003): How can pupils make political judgements? Categories of forming political judgement in the consciousness of pupils, Schwalbach/Ts.

- OECD (Hrsg.) (2002): Lesen kann die Welt verändern. Leistung und Engagement im Ländervergleich. Ergebnisse von PISA 2000, Paris
OECD (publ.) (2002): Reading can change the world. Performance and Engagement. A country comparison. Results of PISA 2000, Paris
- Oesterreich, Detlef (2002): Politische Bildung von 14-Jährigen in Deutschland. Studien aus dem Projekt Civic Education, Opladen
Oesterreich, Detlef (2002): Civic Education of a 14 year-old in Germany. Studies from the project Civic Education, Opladen
- Schelle, Carla (2003): Politisch-historischer Unterricht hermeneutisch rekonstruiert. Von den Ansprüchen Jugendlicher, sich selbst und die Welt zu verstehen, Bad Heilbrunn
Schelle, Carla (2003): Politico-historical teaching reconstructed hermeneutically. The demands of young people to understand themselves and the world, Bad Heilbrunn
- Schwarzer, Christine/Schwarzer, Ralf (1977): Praxis der Schülerbeurteilung. Ein Arbeitsbuch, München
Schwarzer, Christine/Schwarzer, Ralf (1977): The practice of pupil rating. A workbook, Munich
- Stenner, Georg A. (1979): Lernerfolgskontrolle in Sozialkunde, in: Böhmer, Maria (Hrsg.): Lernerfolgskontrolle, Königstein, S. 108 – 133
Stenner, Georg A. (1979): Controls for learning success in civics, in: Boehmer, Maria (publ.): Controls for learning success, Koenigstein, S. 108 – 133
- Richter, Dagmar (2004): Zu Bildungsstandards und möglichen Beiträgen qualitativer Unterrichtsforschung. In: GPJE (Hrsg.): Politische Bildung zwischen individualisiertem Lernen und Bildungsstandards. Schwalbach, S. 119-133
Richter, Dagmar (2004): Educational Standards and possible contributions of qualitative research into teaching. in: GPJE (publ.): Civic Education between individual learning and educational standards. Schwalbach, p. 119-133
- Spitzer, Manfred (2002): Lernen: Gehirnforschung und die Schule des Lebens, Heidelberg/Berlin
Spitzer, Manfred (2002): Learning: Brain research and the school of life, Heidelberg/Berlin
- Weinert, Franz E. (2000): Lehren und Lernen für die Zukunft - Ansprüche an das Lernen in der Schule. URL: http://pz.bildung-rp.de/pn/pn2_00/weinert.htm
Weinert, Franz E. (2000): Teaching and learning for the future – demands on learning in school. URL: http://pz.bildung-rp.de/pn/pn2_00/weinert.htm
- Weißeno, Georg (2004a): Nationale Bildungsstandards für Gemeinschaftskunde – Wohin geht die Bildungspolitik? In: Informationen für den Geschichts- und Gemeinschaftskundelehrer, H. 67, S. 72 – 76
Weißeno, Georg (2004a): National Educational Standards for Social Studies – Where is educational policy heading? In: Information for the history and social studies teacher, Issue. 67, p. 72 – 76
- Weißeno, Georg (2004b): Vom Input zum Output – Zur Zukunft empirischer Untersuchungen für den Politikunterricht, in: Breit, Gotthard/Schiele, Siegfried (Hrsg.): Demokratie braucht politische Bildung, Schwalbach/Ts., S. 149-165
Weißeno, Georg (2004b): From input to output – the future of empirical studies on the teaching of politics, in: Breit, Gotthard/Schiele, Siegfried (publ.): Democracy needs Civic Education, Schwalbach/Ts., p. 149-165
- Weißeno, Georg (2004c): Wie können Lern- und Testaufgaben die Anforderungen der Bildungsstandards für die politische Bildung erfüllen? In: Politische Bildung, Heft 3, S. 70-81
Weißeno, Georg (2004c): How can learning and test tasks fulfil the requirements of the educational standards for Civic Education? In: Civic Education, Issue 3, p. 70-81