

Dealing with flawed items in examinations: Using the compensation of disadvantage as used in German state examinations in items with partial credit scoring

Abstract

In the written part of German state exams, multiple choice questions which inherit a flaw are not always excluded but taken into consideration when awarding a grade to any individual candidate, if this is advantageous to the candidate. This process of elimination and compensation of disadvantage in flawed questions, as applied by the Institut für medizinische und pharmazeutische Prüfungsfragen (IMPP), can lead to *different items* being taken into consideration in the grading *in individual cases* and to the application of *different pass and grade boundaries in individual cases*.

This procedure is applied here to examinations containing items with partial credit scoring. A simple numerical rule can be applied which can be used to decide whether a flawed item has to be taken into account in the grading procedure in order to achieve a pass or a particular grade for a candidate. This rule – as in the state examinations – must take account of the individual number of scores achieved in the flawed item and, if the relative pass threshold ("automatic adjustment clause") is applied, the average score achieved in the item by all candidates. In addition, however, it also includes the grade boundaries.

Keywords: examination, IMPP

1. Background

1.1. Flawed items in examinations

Despite the careful review of examination items *before the examination*, it is almost impossible to avoid the discovery *after the examination* that individual items contain errors of content or form, for example that in a Type A multiple-choice item ("one out of five") in fact two possible answers must be acknowledged as being correct. An item such as this is considered to be "flawed" because, contrary to convention, it permits more than one single correct answer.

Those responsible for examinations within the faculties are therefore faced with the problem of needing to develop a legally sound procedure for dealing with flawed items. Study regulations and examination guidelines frequently contain insufficient directives regarding this aspect, which is why the approach of the Institut für medizinische und pharmazeutische Prüfungsfragen (IMPP), which is seen as legally sound, has been or is soon to be adopted by faculties for their own examinations in some places.

This is possible as long as the examination consists only of items in which no partial scores are awarded, i.e. in which an item can only be scored as having been answered *correctly or incorrectly* by the candidate.

Andreas Möltner¹

1. Medizinische Fakultät
Heidelberg,
Kompetenzzentrum für
Prüfungen in der Medizin,
Heidelberg, Germany

1.2. The elimination procedure and the compensation of disadvantage in flawed items contained in the written component of the state medical examinations

This section presents a short description of the grading method in the written examinations within the German state examination and essentially is a summary and paraphrasing of the explanations given by the IMPP [<https://www.impp.de/internet/de/impp-aktuell.html>], [1].

On the decision regarding whether an examination has been passed and, if yes, how to grade this examination performance, the Ärztliche Approbationsordnung (German medical licensing regulations ÄAppO, section 14 Written Examination) sets out the following [https://www.gesetze-im-internet.de/_appro_2002/]:

(6) *The written examination is deemed to have been passed if the candidate has answered at least 60% of the examination questions posed correctly or if the number of questions answered correctly by the candidate is no more than 22 percent below the average examination performance of the candidates who have participated in the examination for the first time following a minimum study period of two years for the first part the medical examination and five years for the second part of the medical examination.*

(7) Performance in the written examination shall be graded as follows:

If the candidate attains the minimum number of correctly answered examination questions required for passing the examination in accordance with paragraph 6, then he or she shall be awarded the grade

- "very good" if he or she answers at least 75%,
- "good" if he or she answers at least 50% but less than 75%,
- "satisfactory" if he or she answers at least 25% but less than 50%,
- "pass" if he or she answers none or less than 25% of the remaining examination questions correctly.

The grade boundary defined by the 60% criterion shall be known as the "60% threshold". The second criterion ("... 22 percent...") serves to lower the pass threshold in examinations with particularly poor results. For this threshold the term "automatic adjustment clause boundary" (*Gleitklauselgrenze*) is generally used. This is a norm-referenced boundary and depends on the average score achieved in a subgroup of all examination candidates, the "reference group". The reference group is made up of the students who after (precisely) two years or (precisely) five years of study are sitting the examination for the first resp. the second part of the state medical examination for the first time. Consequently the lower of the two thresholds (i. e. the 60% and the automatic adjustment clause threshold) is applied.

Following the examination, all items are once again checked to ensure that they are correct and legitimate on the basis of any appeals/complaints or statistical analyses. During this process it may emerge that an item (i) cannot be correctly answered with the answer options provided, or can be (ii) misinterpreted and/or contain more than one correct answer option. These items are then described as "flawed". The items in group (ii) can thus – whether flawed in content or form – also be answered "correctly".

The flawed questions are "removed from the grading scheme" ("eliminated"); they only count for those candidates who answer the questions justifiably correct (see the judgement of the German Federal Administrative Court of 17.05.1995 [https://www.jurion.de/urteile/bverwg/1995-05-17/bverwg-6-c-8_94/] and the detailed description of elimination and compensation of disadvantage [<https://www.impp.de/internet/de/impp-aktuell.html>], [1]).

In other words: whilst unanswerable flawed questions are in fact completely excluded from the grading (question group i), flawed items which can be answered correctly (question group ii) are only taken into account for those candidates who have actually answered them correctly.

Example 1: If, for example, of the 320 items in the state examination it is established that in two questions two answers could be considered correct then these two questions are eliminated, i.e. the pass threshold is now 60% of 318 items. This calculation results in a value of 190.8. Thus 191 items ("at least 60%") must be answered correctly in order to pass.

If a candidate answers 190 of the 318 properly formulated items correctly and does not give a correct answer in either of the two eliminated items, then he or she has not passed the examination.

If the candidate has chosen a correct answer in one of the two eliminated items, then the score from this answer is awarded to him or her. The candidate has then indeed answered 191 items correctly; at the same time, however, the pass threshold increases to 60% of 319 items and thus a value of 191.4, so that 192 correct answers would have been required in order to pass.

Only if both eliminated items are answered correctly would the example candidate have passed: 192 items have been answered correctly, the pass threshold now stands at 192 for 60% of 320 items.

The result of this is that, depending on how the eliminated items have been answered, different items count towards the grades in individual cases and different pass thresholds may apply in individual cases. Thus, where correct answers are given to flawed ("eliminated") items, a "softening" of the minimum standards set out in the licensing regulations (e.g. at least 60% of questions to be answered correctly) is avoided: for example, if at least two possible answers were correct in 10 of the 320 items, and if a candidate also provided a correct answer in all of these items, then the 60% threshold of the 310 properly formulated items is 186. However, since for this candidate all 320 items are taken into consideration, it is only 186 of 320=58.125%, therefore less than 60%. The pass threshold is therefore defined for this candidate out of the 320 items: 60% of 320=192.

1.3. Partial credit scoring

At many medical faculties, however, in examinations items are also used which do not correspond to the classic "one out of five" format ("Type A"), such as for example multiple true false items (frequently also called "Type X" or "Kprim") or "select n answers" from a list of options ("PickN"), for which *partial credit* is awarded for answers which are partially correct.

In this case, adopting the procedure used in the state examinations is not a trivial matter; for example in an item which has been classified as being "flawed in form" but for which a candidate would receive 0.75 points, the question arises as to whether this item should be taken into consideration when calculating the grade or not.

The question is not only significant because multiple-choice questions with partial credit scoring are already being used at a range of faculties, but also because further developments in skills orientated tests of knowledge (written or computer-based examinations) could also require the use of other types of item in which partial knowledge should also be adequately taken into consideration in the grading. Indeed, faculties are explicitly encouraged to develop examinations further in the Masterplan 2020 [2]. On the other hand, the faculties require a certain legal certitude here which is why they have often

preferred to proceed in a way which is similar to the state examinations.

In the state examinations themselves, the use of items with partial credit scoring is so far apparently not envisaged; for example, on the one hand it is stated in a comment on a draft of the directive on the new regulations in dentistry education ([3], page 165f) that items of this kind "would permit innovative question and answer formats supported by computer alongside the multiple-choice style question in future", and on the other hand it is "not envisaged [...], that a question can have a half-correct answer."

1.4. Objective and Overview

The objective of this paper is to transfer the process of elimination and compensation for disadvantage used in state examinations to examinations which include items with partial credit scoring.

In doing so we shall consider examinations which consist of items in which only non-negative scores values can be achieved and for which the examination result is composed of the addition of the sum of the scores values achieved in the individual items. There is no requirement that all items must be equally weighted; the maximum achievable score in the various items may thus be different.

In Section 2, the definition of pass and grade boundaries in the state examinations will first of all be described and a rule will be formulated from this which can be transferred to examinations containing items with partial credit scoring. A sub-section will then examine the application of rounding in establishing pass and grade boundaries.

In Section 3 there will then follow an explanation of which – in the terminology of the IMPP – "eliminated" items shall be taken into account if the procedure proposed here is applied for the individual candidates. The following sections contain a series of "formulae" which are required for a precise presentation. Readers less familiar with mathematical notation should not be put off, we have tried with a series of calculated examples to make the formulae easy to follow.

2. Pass and Grade Boundaries

2.1. State Examination

In order to describe the regulations of the ÄAppO (in section 14 Written Examination) [https://www.gesetze-im-internet.de/_appro_2002/], the identifiers B_s for the *mathematical pass threshold* according to the 60% rule and B_g for the *mathematical pass threshold* according to the automatic adjustment clause (mean of the average examination performances of the reference group – 22%) shall be used. M stands for the number of items and X_r for the mean of the examination performances achieved by the reference group in those items

(see Tab. 1). With these identifiers the following is true (see also the detailed description in [<https://www.impp.de/internet/de/impp-aktuell.html>], [1])

$$B_s = 0.60 \times M$$

$$B_g = 0.78 \times X_r$$

It should be noted here that the mathematical pass thresholds thus defined are not necessarily whole numbers (in the licensing regulations the term "pass threshold" is *not* used). As set out in [<https://www.impp.de/internet/de/impp-aktuell.html>], the *actual pass threshold* is the smallest whole number which is greater than or equal to the mathematical threshold. If $\text{ceil}(z)$ denotes the rounding up function (also known as the "ceiling function"), then the actual pass threshold is $\text{ceil}(B)$ (for $B=B_s$ or B_g). The mathematical grade boundaries are then

$$(1) N_g = \text{ceil}(B) + g \times (M - \text{ceil}(B))$$

Whereby the boundary between "fail" and "pass" for $g=0$ (pass threshold), the further boundaries for "satisfactory", "good" and "very good" result in $g=0.25, 0.50$ and 0.75 . Note that these boundaries must be *achieved or exceeded* in each case, falling below the boundary even by only a minimal amount results in the poorer grade in each case. The grade boundaries are defined for both B_s and B_g , the lower of the two thresholds in each case is of significance for the students.

2.2. Pass and Grade Boundaries in Items with Partial Credit Scoring

When partial credit scoring is used, it is not possible to define pass and grade boundaries based solely on the number of "correctly answered examination questions". It would be better to speak of "points" achieved, this formulation permits firstly the use of partial points as well as an unequal weighting of items, amongst other things. In most degree or examination regulations, absolute and relative pass thresholds ("automatic adjustment clause") are stipulated. If M describes the maximum number of points achievable (note that in examinations such as the state examinations, in which one point can be achieved for each item, M agrees with the number of items), and X_r describes the mean score achieved in the examination (if a reference group has been defined, use the mean value in this group), then the absolute pass threshold B_s and the relative pass threshold B_g are determined in accordance with the automatic adjustment clause using two constants c_s or c_g .

$$B_s = c_s \times M$$

$$B_g = c_g \times X_r$$

The value of c_s is usually 0.6 in medical examinations ("anyone who achieves 60% of the maximum score passes the examination"), c_g is often 0.78 as in the state examinations ("... anyone achieving not more than 22% ... has failed"), for c_g other values are also used, such as for example at the Medical Faculty of Heidelberg where the value 0.80 is used [4].

The following example shows that the direct application of equation (1) leads to the – in our opinion undesirable

– effect that the additional consideration of items with identical properties may first lead to an improvement and subsequently to a lowering of the grade:

Example 2: let us assume that an examination consists of 26 items in which candidates can achieve points in whole numbers from 0 to 4 points in each case. Two items are taken out of the grading scheme.

A candidate achieved 67 points in the 24 correctly set items. The candidate achieved 3 points in each of the two items removed from the grading scheme.

If only the 24 items included in the grading scheme are taken into consideration and the equation (1) used in the state examinations is directly applied, then for the boundary between "pass" and "satisfactory" the result is as follows (a maximum of $4 \times 24 = 96$ points can be achieved):

$$\begin{aligned} & \text{ceil}(B) + g \times (M - \text{ceil}(B)) \\ &= \text{ceil}(0.60 \times 96) + 0.25 \times (96 - \text{ceil}(0.60 \times 96)) \\ &= \text{ceil}(57.60) + 0.25 \times (96 - \text{ceil}(57.60)) \\ &= 58 + 0.25 \times (96 - 58) = 67.50 \end{aligned}$$

With 67 points the candidate would fall below this boundary, the grade would be "pass".

If we take into account one of the two items removed from the grading scheme, then the result is

$$\begin{aligned} & \text{ceil}(B) + g \times (M - \text{ceil}(B)) \\ &= \text{ceil}(0.60 \times 100) + 0.25 \times (100 - \text{ceil}(0.60 \times 100)) \\ &= \text{ceil}(60.00) + 0.25 \times (100 - \text{ceil}(60.00)) \\ &= 60 + 0.25 \times (100 - 60) = 70.00 \end{aligned}$$

With $67 + 3 = 70$ points of a possible $4 \times 25 = 100$ points, the candidate would receive the grade "satisfactory".

If both of the items taken out of the grading scheme, and in which the candidate achieved 3 out of 4 points each, are taken into account, then

$$\begin{aligned} & \text{ceil}(B) + g \times (M - \text{ceil}(B)) \\ &= \text{ceil}(0.60 \times 104) + 0.25 \times (104 - \text{ceil}(0.60 \times 104)) \\ &= \text{ceil}(62.40) + 0.25 \times (104 - \text{ceil}(62.40)) \\ &= 63 + 0.25 \times (104 - 63) = 73.25 \end{aligned}$$

With $67 + 3 + 3 = 73$ points of a possible $4 \times 26 = 104$ the candidate once again only achieves the grade "pass".

Example 2 shows that if 70% of the maximum number of points is achieved, which is normally associated with the boundary between "pass" and "satisfactory", the addition of one item in which 75% of the number of points (3 out of 4 points) was achieved can imply a downgrading. This seems paradoxical to the author.

Formal Definition of the Pass and Grade Boundaries in Items with Partial Credit Scoring

We can achieve a simpler formal definition than equation (1) for the various grade boundaries associated with the pass thresholds B_s and B_g with

$$(2) N_g = B + g \times (M - B)$$

which distributes the gap between mathematical pass threshold and maximum achievable score into four numerically equal intervals. This is different to equation (1) in that the mathematical pass threshold B and not the rounded up actual pass threshold $\text{ceil}(B)$ is fed into the definition of the grade boundaries.

Using this definition we get around the obvious problem that in an examination with a maximum of 22 achievable points a candidate with 13.5 points would fail, because although the mathematical 60% threshold of $0.6 \times 22 = 13.2$ is exceeded, the candidate would not achieve the rounded up whole number threshold of 14. Further, it is impossible to construct a case similar to example 2 using this definition (the mathematical grade boundaries between "pass" and "satisfactory" calculate out to 67.2, 70.0 and 72.8 using equation (2) for a maximum achievable numbers of points of 96, 100 and 104 points).

The mathematical grade boundaries using equation (2) may be lower than those using equation (1) but never higher. This may cause a discrepancy with the procedure used in the state examination in a particular place. This is conditioned by the formulation "of the remaining examination questions posed" in the *ÄAppO* which means that in equation (1) $\text{ceil}(B)$ must stand instead of B as in equation (2):

Example 3: Let us assume that an examination consists of 317 items. The mathematical pass threshold is calculated at 190.2. In the procedure suggested here, the result for the mathematical grade boundary between satisfactory and pass is the value (see equation 2: $M = 317$, $B_s = 0.6 \times 317 = 190.2$ and with $g = 0.25$)

$$190.2 + 0.25 \times (317 - 190.2) = 221.9$$

If 222 items are answered correctly, the result is "satisfactory". In the state examination the equivalent grade boundary (equation 1) would be

$$\begin{aligned} & \text{ceil}(190.2) + 0.25 \times (317 - \text{ceil}(190.2)) \\ &= 191 + 0.25 \times (317 - 191) = 222.5, \end{aligned}$$

consequently 223 of the items would then have to be answered correctly in order to achieve "satisfactory".

2.3. Rounding

There are, however, justifiable reasons for considering rounding in the definition of pass and grade boundaries: *Example 4: An examination has 17 items in which one whole score can be achieved in each item. According to equation (1) or (2) with the 60% threshold ($0.6 \times 17 = 10.2$) at least 11 items must be answered correctly in order to pass. That is 64.7%, making it significantly more than 60%. If only 10 of the items were required in order to pass, this would be less than 60% at 58.8%, but the deviation from 60% would however be significantly lower (1.2% instead of 4.7%).*

This also occurs in the state examinations, although with significantly lower percentage variations due to the large number of items. In order to reduce the effect of the "tightening up" of the conditions engendered by the requirement for whole numbers in the examination result achieved (see here also the last paragraph in 1.2), it was for this reason suggested in a draft of the changes to the licensing regulations for dentists [3] in section 34 and section 35 that the mathematical pass threshold be rounded down if the first decimal place is 0 to 4, and rounded up if the first decimal place is 5 to 9. The same

applies for the grade boundaries – where the *already rounded pass threshold* is used! These are then described by the equations

$$(3) B^* = \text{floor}(B + \frac{1}{2})$$

$$N_g = \text{floor}(B^* + g \times (M - B^*) + \frac{1}{2})$$

(B^* describes the rounded mathematical pass threshold, $\text{floor}(z)$ is the rounding down function, so that the function $\text{floor}(z + \frac{1}{2})$ rounds down if the first decimal place is between 0 and 4 and up if the first decimal place is between 5 and 9).

The calculation in equation (3) is intended in the draft on the change to the licensing regulations for dentists in practical terms to achieve a lowering of the grade boundaries by half a point in comparison to the current licensing regulations for doctors which uses equation (1). An exception is the boundary point with the precise decimal value of 0.5 which is rounded up.

The effect described in example 2 can also occur with equation (3).

Example 5: The examination consists of 24 items, each with a maximum of 4 points, of which 2 are removed from the grading scheme. The candidate has achieved 3 points in each of these 2 items, in the other 22 items he has achieved 61 points.

Without taking into account the eliminated items, the maximum achievable points for the rounded pass threshold with $4 \times 22 = 88$ are

$$B^* = \text{floor}(B + \frac{1}{2}) = \text{floor}(0.6 \times 88 + 0.5)$$

$$= \text{floor}(52.8 + 0.5) = 53$$

The boundary between "pass" and "satisfactory" using

$$\text{floor}(B^* + g \times (M - B^*) + \frac{1}{2})$$

$$= \text{floor}(53 + 0.25 \times (88 - 53) + 0.5)$$

$$= \text{floor}(53 + 8.75 + 0.5) = 62$$

Taking into account one of the two eliminated items the result, similarly to the boundary between "pass" and "satisfactory", is 64 points and if both of the items are taken into account then it is 68 points.

As in example 2, the sequential consideration of the two eliminated items with identical properties leads first of all to the candidate failing to achieve the boundary between "pass" and "satisfactory" (61 points achieved < 62), then then achieving it (64 points = pass thresholds 64) and then once again failing to achieve it (67 points < 68).

Formal Definition of Rounded Pass and Grade Boundaries in Items with Partial Credit Scoring

In items with partial credit scoring, a rounding of the grade boundaries with a similar effect can be achieved by modifying equation (2), thus:

$$(4) N_g = B + g \times (M - B) - 0.5$$

An analogue to equation (3) is achieved by further requiring that in order to pass, the pass threshold must not only be achieved but exceeded (the same applies for the grade boundaries). As in example 3, however, this does not result in a complete agreement of (3) and (4).

The change from equation (3) to (4) is justified in the same way as that from equation (1) to (2). The desired

reduction by half a point is achieved here by the simple subtraction of the constant 0.5, the requirement that the boundaries are not only achieved but exceeded results from the way in which a case with an exact decimal value of 0.5 is handled (see the last subparagraph).

3. Compensation of Disadvantage in Flawed Items

3.1. State Examination

Section 14 of the licensing regulations for doctors [https://www.gesetze-im-internet.de/_appro_2002/] sets out:

(4) The examination items must [...] be checked as to whether they are flawed, measured against the requirements of paragraph 2, sentence 1. If this check shows that individual examination items are flawed then these must not be taken into account when establishing the examination result. The prescribed number of items for the individual examinations [...] is reduced accordingly. The evaluation of the written examination [...] shall be based on the reduced number of examination items. The reduction in the number of examination items shall not have a negative impact on a candidate.

The last sentence is interpreted in legal terms to mean that flawed items are still to be taken into account in individual cases and will only remain outside of consideration if this is not disadvantageous to the candidate. This means that for each candidate, flawed items are to be considered in such a way that the best possible result for the candidate is achieved, in other words a subset of flawed items is sought which achieves the best possible result for the candidate. This subset is generally not empty.

In the case of the state examinations in which each item can only be marked as correct or incorrect, it can be shown that both for passing and for achieving particular grades, for a particular candidate the best possible result is achieved by the inclusion of all of those eliminated items in which he or she gave a correct answer whilst the other eliminated items are not taken into account.

3.2. Items with Partial Credit Scoring

If the pass and grade boundaries, as defined in sections 2.2 and 2.3 by equation (2) or (4) are applied, it is easy to determine which items should be taken into account. It must be determined whether an eliminated item in which m is the maximum score achievable, x_r points were achieved at the mean (or the mean of the reference group in the item) and p points were achieved by a candidate, should be taken into account for that candidate.

For the grade boundaries derived from B_s , we must also verify here whether

$$(5) g \times m + (1 - g) \times c_s \times m \leq p$$

and for the grade boundaries of the automatic adjustment clause (with BG as the pass threshold)

Table 1: for Example 6: Effect on passing and achieving grade “very good” of taking into account a item with 0.75 points achieved (see text)

Items taken into account	Grade boundary “fail”/“pass”	Total score 1	Grade boundary “good”/“very good”	Total score 2
101	60.60	60.50	90.90	91.00
102	61.20	61.25	91.80	91.75

$$(6) g \times m + (1-g) \times c_s \times x_R \leq p$$

If the inequality condition is met, the item must be taken into account in terms of the respective pass threshold and the grade boundaries derived from it, otherwise not. Note that the items used to determine the performance of the student thus depend on the grade boundary.

Example 6: In an examination, 102 items are set each offering a maximum of 1 point, one of these items is eliminated due to a defect of form. A student achieves 0.75 points in this item. Depending on whether the student lies at the pass threshold or at the “good”/“very good” boundary, taking the eliminated item into account may be favourable or not for the student.

If, for example, he achieves 60.50 points and thus just misses the mathematical pass threshold of 60.60, then taking into account the eliminated item has the result that he now has a score of 61.25 meaning that he now lies above the pass threshold of 61.20 points for 102 items (see Table 1, column “Grade boundary fail/pass” and “total score 1”).

In the inequality (5) we would need to apply $g=0.00$, $m=1$, $c_s=0.60$ and $p=0.75$ and

$$\begin{aligned} g \times m + (1-g) \times c_s \times m \\ = 0.00 \times 1 + (1-0.00) \times 0.6 \times 1 \\ = 0.60 < 0.75 = p \end{aligned}$$

If he has achieved 91.00 points in the marked items, thus gaining “very good” (the boundary is 90.90 points at 101 items) then taking into account the flawed item produces a total score of 91.75 which is not sufficient to achieve the boundary of 91.80 points for a “very good” at 102 items (Table 1, column “grade boundary good/very good” and “total score 2”).

The condition of the inequality (5) is in fact not met here:

$$\begin{aligned} g \times m + (1-g) \times c_s \times m \\ = 0.75 \times 1 + (1-0.75) \times 0.6 \times 1 \\ = 0.90 > 0.75 = p \end{aligned}$$

i.e. the item shall not be included in the grading.

4. Discussion

Adopting the approach to elimination and compensation of disadvantage applied in the state examinations in examinations with unequally weighted items and items with partial credit scoring is not difficult to undertake with a similar definition of the pass and grade boundaries. More minor deviations from the definition of the pass and grade boundaries conditioned by the rounding functions in the setting out of the rules of the licensing regulations (see equation 1 vs. 3) are – as shown in example 3 – required to maintain the consistency of the grading scheme.

For determining those eliminated items which must be individually taken into account for a particular candidate

in order to achieve or exceed pass or grade boundaries, there are simple conditions which are dependent only on the number of points achieved in the item, the grade boundaries and – if the automatic adjustment clause is applied – the mean number of points achieved in the item.

The procedure presented is also suitable for application to grading systems other than a division into four grades once the exam has been passed, which is the focus here. If the values for g in equations (2) or (4) are broken down finely enough, decimal grades can be awarded, for example.

In general, however, the negative consequences of different pass and grade boundaries in individual cases should be considered. The approach described may well be clear in form, but is not always easy for the students to follow. This is the reason, why, for example, in the grading regulations of the Medical Faculty of Heidelberg [4] it has been set out that points achieved in “flawed” items are awarded to the student as *bonus points*; nevertheless, even if bonus points are awarded, only the pass and grade boundaries derived from the correctly formulated items are applied. The “readjustment” of the pass threshold which takes place in the state examinations is not carried out, the same pass and grade boundaries apply for all candidates.

In consideration of the introduction of innovative examination formats promoted in the Masterplan 2020, it would be welcomed if the rules of any future licensing regulations were formulated from the very beginning in such a way that they were *directly* applicable to examinations containing differently weighted items and/or items for which partial scores are awarded (for example, in the state examination for medicine in Switzerland, half points are awarded in multiple true false items [5]). This also applies in the situation that *for the time* in the written state examinations only items for which no partial scores are awarded are used (see [3]), a generally applicable rule would require no further adjustment if such items were *introduced later*. This would equally secure a unified approach for practical examinations (OSCEs, for example) and a simple application of the grading scheme to examinations in which the pass thresholds are determined using the standard setting procedure and further grade boundaries are mathematically derived from it.

Notes

Key to symbols used (see Table 2)

Table 2: Symbols

M	The maximum number of points available in an examination (in examinations in which exactly one point can be achieved in each item this is identical to the number of items).
X_R	Mean value of the number of points achieved in the reference group (if no reference group has been defined, the mean number of points achieved by all candidates).
B_S	Mathematical pass threshold relative to M .
B_G	Mathematical pass threshold relative to X_R (automatic adjustment clause boundary).
B	Mathematical pass threshold (B_S or B_G depending on context).
g	Scaling factor to determine the grade boundary from M and B_S or, if the automatic adjustment clause is applied, from M and B_G . In the German 4-grade system and with equal distribution of the intervals between pass threshold and maximum achievable number of points, for $g = 0$ the result is the pass threshold and for $g = 0.25, 0.50$ and 0.75 the result is the boundaries between "pass" and "satisfactory", "satisfactory" and "good" and "good" and "very good".
c_S	Factor for determining the pass threshold relative to M . In the German state examination $c_S = 0.6$.
c_G	Factor for determining the pass threshold relative to X_R . In the German state examination $c_G = 0.78$.
m	The maximum achievable number of points in a item which is being checked to determine whether the item will be considered in the marking of a candidate's performance or not.
p	The number of points achieved by an examination candidate in this item.
x_R	The mean number of points achieved by the reference group in this item.
ceil(z)	Rounding up function ("ceiling function")
floor(z)	Rounding down function ("floor function")

Funding

This paper was created as part of the MERLIN II project (01PL17011C) funded by the Federal Ministry for Education and Research.

Acknowledgements

I would like to express my particular thanks to Dr. H. Shahla of the Institut für medizinische und pharmazeutische Prüfungsfragen for the intensive and constructive discussions.

Competing interests

The author declares that he has no competing interests.

References

1. Shahla H, Schillings V, Lorenz I. Eliminierung fehlerhafter Prüfungsaufgaben und Nachteilsausgleich als Gebot der Bewertungsfairness. Zur Praxis der Ermittlung individueller Bestehens- und Notengrenzen bei Staatsexamina. 2018. Erhältlich über: hshahla@impp.de
2. Bundesministerium für Bildung und Forschung. Beschlusstext Masterplan Medizinstudium 2020. Berlin: Bundesministerium für Bildung und Forschung; 2017. Zugänglich unter/available from: <https://www.bmbf.de/de/masterplan-medizinstudium-2020-4024.html>
3. Bundesministerium für Gesundheit. Verordnung zur Neuregelung der zahnärztlichen Ausbildung. Berlin: Bundesminister für Gesundheit; 2017. Zugänglich unter/available from <https://www.bundesgesundheitsministerium.de/service/gesetze-und-verordnungen/guv-18-lp.html#c7548>

4. Universität Heidelberg. Studienordnung für das Medizinstudium an der Medizinischen Fakultät Heidelberg der Universität Heidelberg für das 1. und 2. Studienjahr vom 22. Juli 2010. Heidelberg: Universität Heidelberg; 2010. Zugänglich unter/available from: <http://www.uni-heidelberg.de/md/studium/download/03-01-7a-05.pdf>
5. Bundesamt für Gesundheit. Eidgenössische Prüfung in Humanmedizin. Bern: Bundesamt für Gesundheit (BAG). Zugänglich unter/available from: <https://www.bag.admin.ch/bag/de/home/berufe-im-gesundheitswesen/medizinalberufe/eidgenossische-pruefungen-universitaerer-medizinalberufe/eidgenossische-pruefung-in-humanmedizin.html>

Corresponding author:

Dr. Andreas Möltner
Medizinische Fakultät Heidelberg, Kompetenzzentrum für Prüfungen in der Medizin, Im Neuenheimer FGeld 346, D-69120 Heidelberg, Germany, Phone: +49 (0)6221/56-6753
andreas.moeltner@med.uni-heidelberg.de

Please cite as

Möltner A. Dealing with flawed items in examinations: Using the compensation of disadvantage as used in German state examinations in items with partial credit scoring. *GMS J Med Educ*. 2018;35(4):Doc49. DOI: 10.3205/zma001195, URN: urn:nbn:de:0183-zma0011959

This article is freely available from

<http://www.egms.de/en/journals/zma/2018-35/zma001195.shtml>

Received: 2018-03-10

Revised: 2018-07-24

Accepted: 2018-09-25

Published: 2018-11-15

Copyright

©2018 Möltner. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.

Umgang mit fehlerhaften Aufgaben bei Prüfungen: Die Übertragung des Nachteilsausgleichs bei den deutschen Staatsexamina auf Aufgaben mit Teilpunkten

Zusammenfassung

Sind in den schriftlichen Teilen der deutschen Staatsexamina Aufgaben enthalten, die inhaltlich oder formal fehlerhaft sind, werden diese nicht generell ausgeschlossen, sondern bei der Benotung eines einzelnen Prüfungsteilnehmers dennoch berücksichtigt, wenn dies für diesen Teilnehmer *vorteilhaft* ist. Dieses Verfahren von Elimination und Nachteilsausgleich bei fehlerhaften Fragen, wie es vom Institut für medizinische und pharmazeutische Prüfungsfragen (IMPP) angewandt wird, kann dazu führen, dass bei einer Prüfung *individuell unterschiedliche Aufgaben* in der Bewertung berücksichtigt werden sowie *individuell unterschiedliche Bestehens- und Notengrenzen* gelten.

Dieses Vorgehen wird auf Prüfungen übertragen, die Aufgaben enthalten, bei denen Teilpunkte vergeben werden. Es lässt sich ein einfaches numerisches Kriterium angeben, mit dem entschieden werden kann, ob eine fehlerhafte Aufgabe zum Erreichen der Bestehens- oder einer Notengrenze bei einer Prüfungsteilnehmerin oder einem Prüfungsteilnehmer für die Bewertung zu berücksichtigen ist oder nicht. In dieses Kriterium geht – wie bei den Staatsexamina – die individuell erreichte Punktzahl bei der fehlerhaften Aufgabe und bei Anwendung der relativen Bestehensgrenze („Gleitklausel“) die mittlere erreichte Punktzahl aller Teilnehmer bei der Aufgabe ein. Darüber hinaus beinhaltet es aber auch noch die Notengrenzen.

Schlüsselwörter: Prüfung, IMPP

Andreas Möltner¹

1 Medizinische Fakultät
Heidelberg,
Kompetenzzentrum für
Prüfungen in der Medizin,
Heidelberg, Deutschland

1. Hintergrund

1.1. Fehlerhafte Aufgaben in Prüfungen

Es ist nahezu unvermeidlich, dass trotz sorgfältiger Begutachtung von Prüfungsaufgaben *vor der Prüfung*, sich *nach der Prüfung* herausstellt, dass einzelne Aufgaben inhaltlich oder formal fehlerhaft sind, wie z. B., dass in einer Multiple-Choice-Aufgabe vom Typ A („Eins aus Fünf“) tatsächlich zwei Antwortoptionen als zutreffend anerkannt werden müssen. Eine solche Aufgabe gilt als „fehlerhaft“, da sie entgegen der Konvention nicht nur eine einzige korrekte Antwort zulässt.

Prüfungsverantwortliche der Fakultäten stehen damit vor dem Problem, ein juristisch haltbares Vorgehen zum Umgang mit fehlerhaften Aufgaben entwickeln zu müssen. Studien- oder Prüfungsordnungen enthalten hierfür häufig keine ausreichenden Vorgaben, weshalb die als rechtssicher anzusehende Umgangsweise des Instituts für medizinische und pharmazeutische Prüfungsfragen (IMPP) bei den deutschen Staatsexamina mancherorts von den Fakultäten für ihre eigenen Prüfungen übernommen wird oder werden soll.

Dies ist möglich, sofern die Prüfung nur aus Aufgaben besteht, bei denen keine Teilpunkte vergeben werden, d. h. bei denen eine Aufgabe bei der Bewertung nur als entweder vom Prüfungsteilnehmer *zutreffend* beantwortet oder *nicht zutreffend* beantwortet klassifiziert werden kann.

1.2. Das Verfahren von Elimination und Nachteilsausgleich bei fehlerhaften Aufgaben der schriftlichen Teile der medizinischen Staatsexamina

Im diesem Abschnitt soll eine kurze Beschreibung des Bewertungsmodus bei den schriftlichen Prüfungen im deutschen Staatsexamen erfolgen, die im Wesentlichen eine Kürzung und Paraphrasierung der Erläuterungen des IMPP [<https://www.impp.de/internet/de/impp-aktuell.html>], [1] darstellt.

Für die Entscheidung, ob eine Prüfung bestanden ist und, falls ja, wie diese Prüfungsleistung zu benoten ist, gibt die Ärztliche Approbationsordnung (ÄAppO, § 14 Schriftliche Prüfung) Folgendes vor [https://www.gesetze-im-internet.de/_appro_2002/]:

(6) Die schriftliche Prüfung ist bestanden, wenn der Prüfling mindestens 60 Prozent der gestellten Prüfungsfragen zutreffend beantwortet hat oder wenn die Zahl der vom Prüfling zutreffend beantworteten Fragen um nicht mehr als 22 Prozent die durchschnittlichen Prüfungsleistungen der Prüflinge unterschreitet, die nach der Mindeststudienzeit von zwei Jahren beim Ersten Abschnitt der Ärztlichen Prüfung und fünf Jahren beim Zweiten Abschnitt der Ärztlichen Prüfung erstmals an der Prüfung teilgenommen haben.

(7) Die Leistungen in der schriftlichen Prüfung sind wie folgt zu bewerten:

Hat der Prüfling die für das Bestehen der Prüfung nach Absatz 6 erforderliche Mindestzahl zutreffend beantworteter Prüfungsfragen erreicht, so lautet die Note

- „sehr gut“, wenn er mindestens 75 Prozent,
- „gut“, wenn er mindestens 50, aber weniger als 75 Prozent,
- „befriedigend“, wenn er mindestens 25, aber weniger als 50 Prozent,
- „ausreichend“, wenn er keine oder weniger als 25 Prozent

der darüber hinaus gestellten Prüfungsfragen zutreffend beantwortet hat.

Die durch das 60%-Kriterium definierte Grenze sei als „60%-Grenze“ bezeichnet. Das zweite Kriterium („... 22 Prozent ...“) dient dazu, bei besonders schlecht ausgefallenen Prüfungen die Bestehensgrenze abzusenken. Für diese Grenze wird meist die Bezeichnung „Gleitklauselgrenze“ benutzt. Sie ist eine normorientierte Grenze und hängt von der mittleren erreichten Punktzahl einer Untergruppe aller Prüfungsteilnehmer, der „Referenzgruppe“ ab. Als Referenzgruppe werden dabei die Studierenden bezeichnet, die nach (genau) zwei Jahren oder (genau) fünf Jahren Studienzeit zum ersten Mal an der Prüfung zum Ersten bzw. Zweiten Abschnitt der Ärztlichen Prüfung teilnehmen. Zur Anwendung kommt demzufolge die niedrigere der 60%- und der Gleitklauselgrenze.

Nach der Prüfung werden alle Aufgaben noch einmal hinsichtlich ihrer Korrektheit und Zulässigkeit auf Grund eventueller Einsprüche oder statistischen Auffälligkeiten bei der Beantwortung überprüft. Dabei kann sich herausstellen, dass ein Aufgabe mit den vorgegeben Antwortmöglichkeiten (i) nicht zutreffend zu beantworten ist, sie kann auch (ii) missverständlich formuliert sein und/oder mehr als eine zutreffende Antwortoption enthalten. Diese Aufgaben werden als „fehlerhaft“ bezeichnet. Die Aufgaben der Gruppe (ii) können demnach – obwohl inhaltlich oder formal fehlerhaft – auch „zutreffend“ beantwortet werden.

Die fehlerhaften Aufgaben werden „aus der Wertung genommen“ („eliminiert“); sie gelten nur für die Teilnehmer als gestellt, die diese vertretbar zutreffend beantwortet haben (s. hierzu das Urteil des Bundesverwaltungsgerichts vom 17.05.1995 [https://www.jurion.de/urteile/bverwg/1995-05-17/bverwg-6-c-8_94/] und die eingehende Darstellung zu Elimination und Nachteilsausgleich [<https://www.impp.de/internet/de/impp-aktuell.html>], [1]).

Mit anderen Worten: Während völlig unbeantwortbare Fragen tatsächlich vollständig aus der Bewertung ausgeschlossen werden (Fragengruppe i), werden fehlerhafte Aufgaben, die zutreffend beantwortbar sind (Fragengruppe ii) nur bei den Prüfungsteilnehmern berücksichtigt, die sie auch zutreffend beantwortet haben.

Beispiel 1: Wurde z. B. von den 320 Aufgaben im Staatsexamen bei zwei Fragen festgestellt, dass zwei Antworten vertretbar sind, so werden diese beiden Fragen eliminiert, d. h. die Bestehensgrenze ergibt sich aus 60% von 318 Aufgaben. Rechnerisch ergibt sich ein Wert von 190,8. Zum Bestehen sind demnach 191 Aufgaben („mindestens 60%“) korrekt zu beantworten.

Hat ein Prüfungsteilnehmer 190 der 318 fehlerfreien Aufgaben zutreffend beantwortet und bei keiner der beiden eliminierten Aufgaben eine zutreffende Antwort gegeben, so hat er die Prüfung nicht bestanden.

Hat er bei einer der beiden eliminierten Aufgaben eine vertretbare Antwort gewählt, so wird ihm diese Aufgabe angerechnet. Er hat zwar dann 191 Aufgaben korrekt beantwortet, gleichzeitig erhöht sich die Bestehensgrenze aber auf 60% von 319 Aufgaben und somit einen Wert von 191,4, so dass 192 korrekte Antworten zum Bestehen nötig gewesen wären.

Erst wenn beide eliminierten Aufgaben zutreffend beantwortet werden, hätte der Beispielskandidat bestanden: 192 Aufgaben sind korrekt beantwortet, die Bestehensgrenze ist jetzt bei 60% von 320 Aufgaben bei 192.

Dies führt dazu, dass je nach Beantwortung der eliminierten Aufgaben *individuell unterschiedliche Aufgaben in die Bewertung eingehen* sowie *individuell unterschiedliche Bestehensgrenzen* gelten können. Damit wird bei korrekter Beantwortung von fehlerhaften („eliminierten“) Aufgaben eine „Aufweichung“ der in der Approbationsordnung vorgegebenen Mindeststandards (z. B. wenigstens 60% zutreffend zu beantwortende Fragen) vermieden: Waren z. B. bei 10 der 320 Aufgaben mindestens 2 Antwortoptionen zutreffend und hat ein Teilnehmer auch bei all diesen Aufgaben eine korrekte Antwort gegeben, so liegt die 60%-Grenze aus den 310 korrekten Aufgaben bei 186. Da aber für diesen Teilnehmer alle 320 Aufgaben berücksichtigt werden, sind dies lediglich 186 von 320=58,125%, also weniger als 60%. Die Bestehensgrenze wird deshalb *für diesen Teilnehmer* aus den 320 Aufgaben bestimmt: 60% von 320=192.

1.3. Teilpunktvergabe

An vielen medizinischen Fakultäten werden bei den Prüfungen jedoch auch Aufgaben eingesetzt, die nicht dem klassischen „Eins aus Fünf“-Format („Typ A“) entsprechen, wie z. B. Mehrfach-Richtig-Falsch-Aufgaben (häufig auch als „Typ X“ oder „Kprim“ bezeichnet) oder „Wähle n Antworten“ aus einer Liste von Optionen aus („PickN“), für die dann bei teilweise richtige Beantwortung *Teilpunkte* vergeben werden.

In diesem Fall ist eine Übertragung des bei den Staats-examina verwendeten Verfahrens nicht trivial, so stellt sich etwa bei einer Aufgabe, die als „formal fehlerhaft“

eingestuft wird, bei der aber ein Teilnehmer oder eine Teilnehmerin 0,75 Punkte erhalten würde, die Frage, ob diese Aufgabe bei der Notenberechnung berücksichtigt werden soll oder nicht.

Die Frage ist nicht nur deshalb von Bedeutung, als dass an einer Reihe von Fakultäten Multiple-Choice-Aufgaben, bei denen Teilpunkte vergeben werden, bereits Verwendung finden, sondern auch deshalb, weil Weiterentwicklungen bei kompetenzorientierten Wissenstests (schriftliche oder computerbasierte Prüfungen) auch die Verwendung von weiteren Aufgabentypen erforderlich machen können, bei denen auch Teilwissen adäquat in der Bewertung berücksichtigt werden soll. Zur Weiterentwicklung von Prüfungen sind die Fakultäten im Masterplan 2020 [2] ja auch explizit aufgefordert. Andererseits benötigen die Fakultäten dabei auch eine gewisse Rechtssicherheit, weshalb von ihnen bei ihren Prüfungen oft ein Vorgehen präferiert würde, welches analog zu dem der Staatsexamina aufgebaut ist.

Bei den Staatsexamina selbst ist die Verwendung von Aufgaben mit Teilpunkten bislang anscheinend nicht vorgesehen, so findet sich in einem Kommentar eines Entwurfs zur Verordnung zur Neuregelung der zahnärztlichen Ausbildung ([3], S. 165f) einerseits, dass diese „neben dem Antwort-Wahl-Verfahren künftig auch innovative und rechnergestützte Frage- und Antwortformate zu[liefere]“, andererseits sei „nicht vorgesehen [...], dass eine Frage halbrichtig beantwortet werden kann“.

1.4. Ziel und Übersicht

Ziel dieser Arbeit ist, das in den Staatsexamina angewandte Vorgehen von Elimination und Nachteilsausgleich auf Prüfungen zu übertragen, die Aufgaben enthalten, bei denen Teillösungen mit Teilpunkten honoriert werden. Dabei seien Prüfungen betrachtet, die aus Aufgaben bestehen, bei denen nur nicht-negative Punktwerte erreicht werden können und das Prüfungsergebnis sich additiv aus der Summe der bei den einzelnen Aufgaben erreichten Punktwerte zusammensetzt. Es wird nicht vorausgesetzt, dass alle Aufgaben gleich gewichtet sind; die maximal erreichbaren Punktzahlen bei den Aufgaben können also unterschiedlich sein.

Im Abschnitt 2 wird zunächst die Bestimmung von Bestehens- und Notengrenzen bei den Staatsexamina formalisiert und hieraus eine Regelung formuliert, die auf Prüfungen übertragen werden kann, die Aufgaben mit Teilpunkten enthalten. Ein ergänzender Unterabschnitt thematisiert die Anwendung von Rundungen bei der Festlegung von Bestehens- und Notengrenzen.

In Abschnitt 3 erfolgt eine Darstellung, welche – in der Terminologie des IMPP – „eliminierten“ Aufgaben bei Anwendung der hier vorgeschlagenen Verfahren für die individuellen Prüfungsteilnehmer zu berücksichtigen sind. Die folgenden Abschnitte enthalten eine Reihe von „Formeln“, die für eine exakte Darstellung erforderlich sind. Die mit mathematischen Notationen weniger vertrauten Leser mögen sich dadurch nicht abschrecken lassen, es

wurde mit einer Reihe von gerechneten Beispielen versucht, diese nachvollziehbar zu machen.

2. Bestehens- und Notengrenzen

2.1. Staatsexamen

Zur Formalisierung der Vorgaben der ÄAppO (in § 14 Schriftliche Prüfung) [https://www.gesetze-im-internet.de/_appro_2002/] seien die Bezeichnungen B_s für die rechnerische Bestehensgrenze nach der 60%-Regel und B_g für die rechnerische Bestehensgrenze nach der Gleitklausel (Mittel der durchschnittlichen Prüfungsleistungen der Referenzgruppe – 22%) verwendet. M steht für die Anzahl der Aufgaben und X_r für das Mittel der von der Referenzgruppe dabei erreichten Prüfungsleistungen (siehe Tab. 1). Mit diesen Bezeichnungen gilt (siehe auch die detaillierte Beschreibung in [<https://www.impp.de/internet/de/impp-aktuell.html>], [1])

$$B_s = 0,60 \times M$$

$$B_g = 0,78 \times X_r$$

Man beachte dabei, dass die so definierten rechnerischen Bestehensgrenzen nicht notwendigerweise ganzzahlig sind (in der Approbationsordnung wird der Terminus „Bestehensgrenze“ *nicht* verwendet). Wie in [<https://www.impp.de/internet/de/impp-aktuell.html>] ausgeführt, ist die *faktische Bestehensgrenze* die kleinste ganze Zahl, die größer oder gleich der rechnerischen Grenze ist. Bezeichnet $\text{ceil}(z)$ die Aufrundungsfunktion, so ist die faktische Bestehensgrenze $\text{ceil}(B)$ (für $B=B_s$ oder B_g). Die rechnerischen Notengrenzen ergeben sich dann zu

$$(1) N_g = \text{ceil}(B) + g \times (M - \text{ceil}(B))$$

wobei die Grenze zwischen „nicht bestanden“ und „ausreichend“ für $g=0$ gegeben ist (Bestehensgrenze), die weiteren Grenzen für „befriedigend“, „gut“ und „sehr gut“ erhält man für $g=0,25, 0.50$ bzw. 0.75 . Zu beachten ist, dass diese Grenzen jeweils *erreicht oder überschritten* werden müssen, eine auch nur minimale Unterschreitung führt zur jeweils schlechteren Note. Die Notengrenzen werden sowohl für B_s wie auch B_g bestimmt, maßgeblich für den Studierenden ist die jeweils Niedrigere der beiden Grenzen.

2.2. Bestehens- und Notengrenzen bei Aufgaben mit Teilpunkten

Bei Verwendung von Teilpunkten ist eine Definition von Bestehens- und Notengrenzen, die allein auf der Zahl der „zutreffend beantworteten Prüfungsfragen“ fußt, nicht anwendbar. Es sollte deshalb besser von erreichten „Punkten“ ausgegangen werden, diese Formulierung erlaubt zum einen die Verwendung von Teilpunkten wie auch eine u. U. ungleiche Gewichtung von Aufgaben. Absolute und relative Bestehensgrenzen („Gleitklausel“) sind in den meisten Studien- oder Prüfungsordnungen vorgegeben. Bezeichnet M die maximal zu erreichende Punktzahl (man beachte, dass bei Prüfungen, wie etwa den Staatsexamina, in denen je Aufgabe genau ein Punkt

erreichbar ist, M mit der Zahl der Aufgaben übereinstimmt), X_r das Mittel der bei der Prüfung erreichten Punktzahl (bei Definition einer Referenzgruppe ist deren Mittel zu verwenden), so sind die absolute Bestehensgrenze B_s und die relative Bestehensgrenze B_g entsprechend der Gleitklausel mittels zweier Konstanten c_s bzw. c_g festgelegt:

$$B_s = c_s \times M$$

$$B_g = c_g \times X_r$$

Der Wert von c_s ist bei medizinischen Prüfungen meist 0.6 („bestanden hat, wer 60% der maximal erreichbaren Punktzahl erreicht hat“), c_g ist oft wie bei den Staatsexamina 0.78 („... wer nicht mehr als 22% ... unterschreitet“), für c_g werden mitunter auch andere Werte verwendet, wie z. B. an der Medizinischen Fakultät Heidelberg der Wert 0.80 [4].

Das nachfolgende Beispiel zeigt, dass die direkte Übertragung von Gleichung (1) zu dem – unseres Erachtens unerwünschten – Effekt führt, dass die zusätzliche Berücksichtigung von Aufgaben mit identischen Eigenschaften erst zu einer Verbesserung und danach wieder zu einer Verschlechterung der Note führen kann:

Beispiel 2: Eine Prüfung bestehe aus 26 Aufgaben, bei denen jeweils ganzzahlige Punktwerte von 0 bis 4 Punkten erreicht werden können. Zwei Aufgaben seien aus der Wertung genommen.

Ein Prüfungsteilnehmer habe bei den 24 korrekten Aufgaben 67 Punkte erreicht, bei den beiden aus der Wertung genommenen Aufgaben jeweils 3 Punkte.

Werden nur die 24 gewerteten Aufgaben berücksichtigt und die bei den Staatsexamina verwendete Gleichung (1) direkt verwendet, ergibt sich für die Grenze zwischen „ausreichend“ und „befriedigend“ (maximal können $4 \times 24 = 96$ Punkte erreicht werden):

$$\begin{aligned} & \text{ceil}(B) + g \times (M - \text{ceil}(B)) \\ &= \text{ceil}(0,60 \times 96) + 0,25 \times (96 - \text{ceil}(0,60 \times 96)) \\ &= \text{ceil}(57,60) + 0,25 \times (96 - \text{ceil}(57,60)) \\ &= 58 + 0,25 \times (96 - 58) = 67,50 \end{aligned}$$

Mit 67 Punkten läge man unter dieser Grenze, die Note wäre „ausreichend“.

Berücksichtigt man eine der beiden aus der Wertung genommenen Aufgaben, gilt

$$\begin{aligned} & \text{ceil}(B) + g \times (M - \text{ceil}(B)) \\ &= \text{ceil}(0,60 \times 100) + 0,25 \times (100 - \text{ceil}(0,60 \times 100)) \\ &= \text{ceil}(60,00) + 0,25 \times (100 - \text{ceil}(60,00)) \\ &= 60 + 0,25 \times (100 - 60) = 70,00 \end{aligned}$$

Mit $67 + 3 = 70$ Punkten von möglichen $4 \times 25 = 100$ Punkten, hätte der Prüfling die Note „befriedigend“.

Werden beide Aufgaben berücksichtigt, die aus der Wertung genommen wurden und bei denen jeweils 3 von 4 Punkten erreicht wurden, so ist

$$\begin{aligned} & \text{ceil}(B) + g \times (M - \text{ceil}(B)) \\ &= \text{ceil}(0,60 \times 104) + 0,25 \times (104 - \text{ceil}(0,60 \times 104)) \\ &= \text{ceil}(62,40) + 0,25 \times (104 - \text{ceil}(62,40)) \\ &= 63 + 0,25 \times (104 - 63) = 73,25 \end{aligned}$$

Mit $67 + 3 + 3 = 73$ Punkten von möglichen $4 \times 26 = 104$ ergibt sich wieder lediglich die Note „ausreichend“.

Das Beispiel 2 zeigt, dass bei Erreichen von 70% der maximalen Punktzahl, die normalerweise mit der Grenze

von „ausreichend“ zu „befriedigend“ assoziiert ist, die Hinzunahme einer Aufgabe, bei der 75% der Punktzahl (3 von 4 Punkten) erzielt wurden, mit einer Verschlechterung der Note einhergehen kann. Dies erscheint dem Autor als widersinnig.

Formale Definition der Bestehens- und Notengrenzen bei Aufgaben mit Teilpunkten

Eine formal einfachere Bedingung als Gleichung (1) für die den verschiedenen Bestehensgrenzen B_s und B_g zugehörigen Notengrenzen kann mit

$$(2) N_g = B + g \times (M - B)$$

erreicht werden, durch die die Spannweite zwischen *rechnerischer Bestehensgrenze und maximal erreichbarer Punktzahl* in vier numerisch gleich große Intervalle unterteilt wird. Der Unterschied zu Gleichung (1) besteht darin, dass in die Bestimmung der Notengrenzen die rechnerische Bestehensgrenze B und nicht die aufgerundete faktische Bestehensgrenze $\text{ceil}(B)$ einfließt.

Durch diese Definition umgeht man das offensichtliche Problem, dass bei einer Prüfung mit maximal 22 erreichbaren Punkten ein Teilnehmer mit 13,5 Punkten durchgefallen wäre, da zwar die rechnerische 60%-Grenze von $0,6 \times 22 = 13,2$ überschritten, die auf eine ganze Zahl aufgerundete Grenze von 14 aber nicht erreicht wäre. Weiter kann damit kein dem Beispiel 2 analoger Fall konstruiert werden (die rechnerischen Notengrenzen zwischen „ausreichend“ und „befriedigend“ errechnen sich nach Gleichung 2 zu 67,2, 70,0 und 72,8 für maximal erreichbare Punktzahlen von 96, 100 bzw. 104 Punkten).

Die rechnerischen Notengrenzen nach Gleichung (2) können niedriger als die nach Gleichung (1) sein, niemals aber höher. Dadurch kann an einer Stelle eine Diskrepanz zum Vorgehen des Staatsexamens auftreten. Dies ist durch die Formulierung „der darüber hinaus gestellten Prüfungsfragen“ in der ÄAppO bedingt, durch die in der Gleichung (1) $\text{ceil}(B)$ statt wie in Gleichung (2) B stehen muss:

Beispiel 3: Eine Prüfung bestehe aus 317 Aufgaben. Die rechnerische Bestehensgrenze errechnet sich zu 190,2. Beim hier vorgeschlagenen Vorgehen ergibt sich für die rechnerische Notengrenze zwischen 3 und 4 der Wert (siehe Gleichung 2: $M = 317$, $B_s = 0,6 \times 317 = 190,2$ und mit $g = 0,25$)

$$190,2 + 0,25 \times (317 - 190,2) = 221,9$$

Mit 222 zutreffend beantworteten Aufgaben erhält man die Note 3. Im Staatsexamen wäre die entsprechende Notengrenze (Gleichung 1)

$$\begin{aligned} & \text{ceil}(190,2) + 0,25 \times (317 - \text{ceil}(190,2)) \\ &= 191 + 0,25 \times (317 - 191) = 222,5, \end{aligned}$$

mithin müssten dann 223 Aufgaben zutreffend beantwortet werden, um „befriedigend“ (Note 3) zu erhalten.

2.3. Rundungen

Es gibt jedoch nachvollziehbare Gründe, Rundungen bei der Bestimmung von Bestehens- und Notengrenzen in Betracht zu ziehen:

Beispiel 4: Eine Prüfung hat 17 Aufgaben, bei denen jeweils ein ganzer Punkt erreicht werden kann. Zum Bestehen sind nach Gleichung (1) oder (2) mit der 60%-Grenze ($0,6 \times 17 = 10,2$) mindestens 11 Aufgaben korrekt zu beantworten, um zu bestehen. Das sind 64,7%, also deutlich mehr als 60%. Würden lediglich 10 Aufgaben zum Bestehen erforderlich sein, wären dies mit 58,8% zwar weniger als 60%, die Abweichung zu 60% wäre aber deutlich geringer (1,2% statt 4,7%).

Auch bei den Staatsexamina tritt dies auf, wenn auch wegen der großen Zahl von Aufgaben mit deutlich geringeren prozentualen Abweichungen. Um die durch die Ganzzahligkeit des erreichten Prüfungsergebnisses bedingte „Verschärfung“ der Bedingungen zu vermindern (vgl. hierzu auch den letzten Absatz in 1.2), wurde deshalb in einem Entwurf zur Änderung der Approbationsordnung für Zahnärzte [3] in § 34 und § 35 vorgeschlagen, die rechnerische Bestehensgrenze bei einer ersten Nachkommastelle von 0 bis 4 ab-, und von 5 bis 9 aufzurunden. Gleiches gilt für die Notengrenzen – bei Verwendung der bereits gerundeten Bestehensgrenze! –, so dass diese durch die Gleichungen

$$(3) B^* = \text{floor}(B + \frac{1}{2})$$

$$N_g = \text{floor}(B^* + g \times (M - B^*) + \frac{1}{2})$$

formalisiert sind (B^* bezeichne die gerundete rechnerische Bestehensgrenze, $\text{floor}(z)$ ist die Abrundungsfunktion, somit rundet die Funktion $\text{floor}(z + \frac{1}{2})$ bei erster Nachkommastelle zwischen 0 und 4 ab und zwischen 5 und 9 auf). Durch die Rundung in Gleichung (3) soll in dem Entwurf zur Änderung der Approbationsordnung für Zahnärzte praktisch eine Absenkung der Notengrenzen um einen halben Punkt gegenüber der gegenwärtig gültigen Approbationsordnung für Ärzte nach Gleichung (1) erzielt werden. Eine Ausnahme ist der Grenzpunkt mit dem exakten Nachkommateil 0,5, der aufgerundet wird.

Auch mit Gleichung (3) kann der in Beispiel 2 beschriebene Effekt auftreten.

Beispiel 5: Die Prüfung bestehe aus 24 Aufgaben mit jeweils maximal 4 Punkten, von denen zwei aus der Wertung genommen sind. Der Prüfungsteilnehmer hat bei diesen je 3 Punkte erreicht, bei den anderen 22 Aufgaben 61 Punkte.

Ohne Berücksichtigung der eliminierten Aufgaben ergibt sich für die gerundete Bestehensgrenze bei $4 \times 22 = 88$ maximal erreichbaren Punkten

$$B^* = \text{floor}(B + \frac{1}{2})$$

$$= \text{floor}(0,6 \times 88 + 0,5)$$

$$= \text{floor}(52,8 + 0,5) = 53$$

Die Grenze zwischen „ausreichend“ und „befriedigend“ liegt bei

$$\text{floor}(B^* + g \times (M - B^*) + \frac{1}{2})$$

$$= \text{floor}(53 + 0,25 \times (88 - 53) + 0,5)$$

$$= \text{floor}(53 + 8,75 + 0,5) = 62$$

Bei Berücksichtigung von einer der beiden eliminierten Aufgaben erhält man analog für die Grenze zwischen „ausreichend“ und „befriedigend“ 64 Punkte und bei Berücksichtigung beider Aufgaben 68 Punkte.

Wie im Beispiel 2 führt die sequentielle Berücksichtigung der beiden eliminierten Aufgaben mit identischen Eigen-

schaften dazu, dass ausgehend davon, dass zunächst die Grenze zwischen „ausreichend“ und „befriedigend“ unterschritten ist (61 erreichte Punkte < 62), dann erreicht (64 Punkte = Bestehensgrenze 64) und schließlich wieder unterschritten wird (67 Punkte < 68).

Formale Definition gerundeter Bestehens- und Notengrenzen bei Aufgaben mit Teilpunkten

Bei Aufgaben mit Teilpunkten kann einer Rundung der Notengrenzen ähnlicher Effekt durch Modifikation von Gleichung (2) mit

$$(4) N_g = B + g \times (M - B) - 0,5$$

erreicht werden. Ein Analogon zu Gleichung (3) erhält man durch die weitere Festlegung, dass zum Bestehen nicht nur das Erreichen sondern ein Überschreiten der Bestehensgrenze erforderlich ist (gleiches gilt für die Notengrenzen). Wie in Beispiel 3 besteht damit aber keine völlige Übereinstimmung von (3) und (4).

Der Übergang von Gleichung (3) zu (4) begründet sich analog zu dem von Gleichung (1) zu (2). Die angestrebte Absenkung um einen halben Punkt wird hier einfach durch die Subtraktion des konstanten Werts 0,5 erreicht, die Notwendigkeit, dass die Grenzen nicht nur erreicht sondern überschritten werden müssen, ergibt sich aus der Behandlung des Falles mit exaktem Nachkommateil 0,5 (siehe den letzten Unterabschnitt).

3. Nachteilsausgleich bei fehlerhaften Aufgaben

3.1. Staatsexamen

Die Approbationsordnung für Ärzte [https://www.gesetze-im-internet.de/_appro_2002/] schreibt in §14 vor:

(4) Die Prüfungsaufgaben sind [...] darauf zu überprüfen, ob sie, gemessen an den Anforderungen des Absatzes 2 Satz 1, fehlerhaft sind. Ergibt diese Überprüfung, dass einzelne Prüfungsaufgaben fehlerhaft sind, sind diese bei der Feststellung des Prüfungsergebnisses nicht zu berücksichtigen. Die vorgeschriebene Zahl der Aufgaben für die einzelnen Prüfungen [...] mindert sich entsprechend. Bei der Bewertung der schriftlichen Prüfung [...] ist von der verminderten Zahl der Prüfungsaufgaben auszugehen. Die Verminderung der Zahl der Prüfungsaufgaben darf sich nicht zum Nachteil eines Prüflings auswirken.

Der letzte Satz wird juristisch so interpretiert, dass fehlerhafte Aufgaben individuell doch zu berücksichtigen sind und nur insoweit unberücksichtigt bleiben, sofern dies nicht nachteilig für den Teilnehmer ist. Das bedeutet, dass für jeden Prüfungsteilnehmer fehlerhafte Aufgaben so zu berücksichtigen sind, dass ein für den Teilnehmer optimales Ergebnis erzielt wird, m. a. W. es wird eine Teilmenge von fehlerhaften Aufgaben gesucht, bei der

Tabelle 1: zu Beispiel 6: Effekt der Berücksichtigung einer Aufgabe mit 0,75 erreichten Punkten auf Bestehen und Erreichen der Note 1 (siehe Text)

Berücksichtigte Aufgaben	Bestehens-Grenze	Erreichte Punktzahl 1	Notengrenze zu Note 1	Erreichte Punktzahl 2
101	60.60	60.50	90.90	91.00
102	61.20	61.25	91.80	91.75

das für den Teilnehmer bestmögliche Resultat erzielt wird. Diese Teilmenge ist i. A. nicht eindeutig.

Bei den Staatsexamina, bei denen jede Aufgabe nur als zutreffend oder nicht zutreffend beantwortet gewertet werden kann, lässt sich zeigen, dass sowohl für Bestehen als auch das Erreichen von Noten die Einbeziehung all derjenigen eliminierten Aufgaben für einen Studierenden zu einem für diesen optimalen Ergebnis führt, bei denen er eine zutreffende Antwort gegeben hat, während die anderen eliminierten Aufgaben unberücksichtigt bleiben.

3.2. Aufgaben mit Teilpunkten

Bei Verwendung der Bestehens- und Notengrenzen, wie sie in den Abschnitten 2.2 und 2.3 durch Gleichung (2) oder (4) definiert sind, können die zu berücksichtigenden Aufgaben einfach festgestellt werden.

Zu bestimmen ist, ob eine eliminierten Aufgabe, bei der maximal m Punkte erreichbar sind, im Mittel x_r Punkte erzielt (evtl. Mittel der Referenzgruppe bei der Aufgabe) und von einem Teilnehmer p Punkte erreicht wurden, für diesen Teilnehmer berücksichtigt werden soll.

Für die aus B_s abgeleiteten Notengrenzen prüft man hierzu, ob

$$(5) \quad g \times m + (1-g) \times c_s \times m < p$$

(analog mit c_a statt c_s für aus B_a abgeleitete Notengrenzen) und für die Notengrenzen der Gleitklausel (mit B_g als Bestehensgrenze)

$$(6) \quad g \times m + (1-g) \times c_g \times x_r \leq p$$

Falls ja, ist die Aufgabe bezüglich der entsprechenden Bestehens- und den daraus abgeleiteten Notengrenzen zu berücksichtigen, ansonsten nicht.

Man beachte, dass die Aufgaben, die zur Bestimmung der Leistung des Studierenden herangezogen werden, damit abhängig von der Notengrenze sind.

Beispiel 6: Seien in einer Prüfung 102 Aufgaben mit jeweils maximal erreichbarer Punktzahl 1 gegeben, von denen eine aus formalen Gründen eliminiert ist. Ein Studierender habe bei dieser Aufgabe 0,75 Punkte erreicht. Abhängig davon, ob der Studierende bei den gewerteten Aufgaben an der Bestehensgrenze oder der Grenze zur Note 1 liegt, kann die Berücksichtigung der eliminierten Aufgabe für den Studierenden nützlich oder schädlich sein.

Hat er etwa mit 60,50 erreichten Punkten die rechnerische Bestehensgrenze 60,60 knapp verfehlt, führt die Berücksichtigung der eliminierten Aufgaben dazu, dass er mit 61,25 Punkten oberhalb der Bestehensgrenze von 61,20 Punkten für 102 Aufgaben liegt (siehe Tabelle 1, Spalte „Bestehensgrenze“ und „Erreichte Punktzahl 1“). In der Ungleichung (5) wäre $g=0,00$, $m=1$, $c_s=0,60$ und $p=0,75$ einzusetzen, es ist

$$g \times m + (1-g) \times c_s \times m$$

$$= 0,00 \times 1 + (1-0,00) \times 0,6 \times 1$$

$$= 0,60 \leq 0,75 = p$$

Damit ist das Kriterium (5) zur Aufnahme in die Bewertung der Aufgabe für diesen Prüfungsteilnehmer erfüllt. Hat er bei den gewerteten Aufgaben 91,00 Punkte erreicht und damit die Note 1 (Grenze ist bei 101 Aufgaben bei 90,90 Punkten) führt die Berücksichtigung der Aufgabe zu einer Punktzahl von 91,75 Punkten, die nicht genügt, die Grenze von 91,80 Punkten zur Note 1 bei 102 Aufgaben zu erreichen (siehe Tabelle 1, Spalte „Notengrenze zu Note 1“ und „Erreichte Punktzahl 2“). Die Bedingung der Ungleichung (5) ist in der Tat hier nicht erfüllt:

$$g \times m + (1-g) \times c_s \times m$$

$$= 0,75 \times 1 + (1-0,75) \times 0,6 \times 1$$

$$= 0,90 > 0,75 = p$$

d. h. die Aufgabe ist nicht zur Bewertung heranzuziehen.

4. Diskussion

Die Übertragung der Vorgehensweise zu Elimination und Nachteilsausgleich, wie sie bei den Staatsexamina verwendet wird, lässt sich durch eine analoge Definition der Bestehens- und Notengrenzen auf Prüfungen mit ungleich gewichteten Aufgaben und Aufgaben mit Teilbewertung unschwer vornehmen. Kleinere Abweichungen von der Definition der Bestehens- und Notengrenzen, die durch die Rundungsfunktionen in der Formalisierung der Regeln der Approbationsordnung bedingt sind (siehe Gleichung 1 vs. 3) sind – wie in Beispiel 3 dargestellt – zur Wahrung der Konsistenz des Notengebungsschemas erforderlich. Für die Bestimmung derjenigen eliminierten Aufgaben, die individuell für einen Prüfungsteilnehmer herangezogen werden müssen, um Bestehens- oder Notengrenzen zu erreichen oder zu überschreiten, existieren einfache Bedingungen, die allein von der bei der Aufgabe erzielten Punktzahl, der Notengrenze und – bei der evtl. notwendigen Anwendung der Gleitklausel – im Mittel erzielten Punktzahl der Aufgabe abhängig ist.

Das dargestellte Vorgehen lässt sich auch auf andere Benotungssysteme als die hier im Fokus stehende Aufteilung in vier Notenstufen bei Bestehen übertragen. Bei entsprechend feinerer Aufteilung der Werte für g in den Gleichungen (2) oder (4) lassen sich z. B. Dezimalnoten vergeben.

Generell sollten aber auch die nachteiligen Konsequenzen individuell unterschiedlicher Bestehens- und Notengrenzen bedacht werden. Das beschriebene Vorgehen ist zwar formal eindeutig, für die Studierenden aber nicht immer leicht nachvollziehbar. Aus diesem Grund wurde z. B. in der Studienordnung der medizinischen Fakultät Heidelberg [4] festgelegt, dass die bei „fehlerhaften“ Aufgaben

Tabelle 2: verwendete Symbole

M	Maximal erreichbare Punktzahl in einer Prüfung (bei Prüfungen, in denen je Aufgabe genau ein Punkt erreicht werden kann, ist diese identisch mit der Zahl der Aufgaben).
X_R	Mittelwert der erreichten Punktzahl in der Referenzgruppe (falls keine Referenzgruppe definiert ist, die mittlere erreichte Punktzahl aller Teilnehmer).
B_S	Rechnerische Bestehensgrenze relativ zu M .
B_G	Rechnerische Bestehensgrenze relativ zu X_R (Gleitklauselgrenze).
B	Rechnerische Bestehensgrenze (je nach Kontext B_S oder B_G).
g	Skalierungsfaktor zur Bestimmung der Notengrenzen aus M und B_S oder bei Anwendung der Gleitklausel aus M und B_G . Beim deutschen Notensystem mit 4 Stufen und gleichmäßiger Aufteilung des Intervalls zwischen Bestehensgrenze und maximal erreichbarer Punktzahl erhält man für $g = 0$ die Bestehensgrenze und für $g = 0,25, 0,50$ und $0,75$ die Grenzen zwischen „ausreichend“ und „befriedigend“, „befriedigend“ und „gut“ bzw. „gut“ und „sehr gut“.
c_S	Faktor zur Bestimmung der Bestehensgrenze relativ zu M . Beim deutschen Staatsexamen ist $c_S = 0,6$.
c_G	Faktor zur Bestimmung der Bestehensgrenze relativ zu X_R . Beim deutschen Staatsexamen ist $c_G = 0,78$.
m	Maximal erreichbare Punktzahl einer Aufgabe, für die geprüft werden soll, ob sie zur Bewertung eines Prüfungsteilnehmers herangezogen werden soll.
p	Von einem Prüfungsteilnehmer bei dieser Aufgabe erreichte Punktzahl.
X_R	Von der Referenzgruppe bei dieser Aufgabe mittlere erreichte Punktzahl.
$\text{ceil}(z)$	Aufrundungsfunktion
$\text{floor}(z)$	Abrundungsfunktion

erzielten Punkte dem Studierenden als *Bonuspunkte* angerechnet werden, jedoch auch bei diesen allein die aus den fehlerfreien Aufgaben bestimmten Bestehens- und Notengrenzen zur Anwendung kommen. Das bei den Staatsexamina erfolgende „Nachjustieren“ der Bestehensgrenze unterbleibt, für alle Teilnehmer gelten die gleichen Bestehens- und Notengrenzen.

In Anbetracht der im Masterplan 2020 geforderten Einführung innovativer Prüfungsformate wäre es zu begrüßen, wenn die Regeln einer zukünftigen Approbationsordnung von vorneherein so gefasst würden, dass sie *direkt* auf Prüfungen, die unterschiedlich gewichtete Aufgaben und/oder Aufgaben mit Teilpunktvergabe enthalten, anwendbar wären (z. B. werden in der Staatsprüfung der Schweiz für Humanmedizin bei Mehrfach-Richtig-Falsch-Aufgaben halbe Punkte vergeben [5]). Dies gilt auch für den Fall, dass vorerst bei den schriftlichen Staatsexamina nur solche Aufgaben eingesetzt werden sollen, für die keine Teilpunkte vergeben werden (vgl. [3]), eine allgemein anwendbare Regel würde bei einer eventuell *späteren Einführung* solcher Aufgaben keine neuerliche Anpassung erfordern. Ebenfalls gesichert wäre damit ein einheitliches Vorgehen für praktische Prüfungen (etwa OSCEs) und eine einfache Übertragung des Schemas der Notengebung auf Prüfungen, in denen die Bestehensgrenzen durch Standard-Setting-Verfahren ermittelt und die weiteren Notengrenzen rechnerisch daraus abgeleitet werden.

Anmerkung

Verzeichnis der verwendeten Symbole (siehe Tabelle 2)

Förderung

Die Arbeit entstand im Rahmen des vom Bundesministerium für Bildung und Forschung geförderten Projekts MERLIN II (01PL17011C).

Danksagung

Besonderen Dank für intensive und konstruktive Diskussionen möchte ich Herrn Dr. H. Shahla vom Institut für medizinische und pharmazeutische Prüfungsfragen aussprechen.

Interessenkonflikt

Der Autor erklärt, dass er keine Interessenkonflikte im Zusammenhang mit diesem Artikel hat.

Literatur

- Shahla H, Schillings V, Lorenz I. Eliminierung fehlerhafter Prüfungsaufgaben und Nachteilsausgleich als Gebot der Bewertungsfairness. Zur Praxis der Ermittlung individueller Bestehens- und Notengrenzen bei Staatsexamina. 2018. Erhältlich über: hshahla@impp.de
- Bundesministerium für Bildung und Forschung. Beschlusstext Masterplan Medizinstudium 2020. Berlin: Bundesministerium für Bildung und Forschung; 2017. Zugänglich unter/available from: <https://www.bmbf.de/de/masterplan-medizinstudium-2020-4024.html>
- Bundesministerium für Gesundheit. Verordnung zur Neuregelung der zahnärztlichen Ausbildung. Berlin: Bundesminister für Gesundheit; 2017. Zugänglich unter/available from <https://www.bundesgesundheitsministerium.de/service/gesetz-und-verordnungen/guv-18-lp.html#c7548>
- Universität Heidelberg. Studienordnung für das Medizinstudium an der Medizinischen Fakultät Heidelberg der Universität Heidelberg für das 1. und 2. Studienjahr vom 22. Juli 2010. Heidelberg: Universität Heidelberg; 2010. Zugänglich unter/available from: <http://www.uni-heidelberg.de/md/studium/download/03-01-7a-05.pdf>
- Bundesamt für Gesundheit. Eidgenössische Prüfung in Humanmedizin. Bern: Bundesamt für Gesundheit (BAG). Zugänglich unter/available from: <https://www.bag.admin.ch/bag/de/home/berufe-im-gesundheitswesen/medizinalberufe/eidgenoessische-pruefungen-universitaerer-medizinalberufe/eidgenoessische-pruefung-in-humanmedizin.html>

Korrespondenzadresse:

Dr. Andreas Möltner
Medizinische Fakultät Heidelberg, Kompetenzzentrum
für Prüfungen in der Medizin, Im Neuenheimer Feld 346,
69120 Heidelberg, Deutschland, Tel.: +49
(0)6221/56-6753
andreas.moeltner@med.uni-heidelberg.de

Bitte zitieren als

Möltner A. Dealing with flawed items in examinations: Using the
compensation of disadvantage as used in German state examinations
in items with partial credit scoring. *GMS J Med Educ.*
2018;35(4):Doc49.
DOI: 10.3205/zma001195, URN: urn:nbn:de:0183-zma0011959

Artikel online frei zugänglich unter

<http://www.egms.de/en/journals/zma/2018-35/zma001195.shtml>

Eingereicht: 10.03.2018

Überarbeitet: 24.07.2018

Angenommen: 25.09.2018

Veröffentlicht: 15.11.2018

Copyright

©2018 Möltner. Dieser Artikel ist ein Open-Access-Artikel und steht
unter den Lizenzbedingungen der Creative Commons Attribution 4.0
License (Namensnennung). Lizenz-Angaben siehe
<http://creativecommons.org/licenses/by/4.0/>.