

The tension between supervision and evaluation: Are tutors able to objectively evaluate an undergraduate dissertation?

Author: Olaf Spittaels

Institution: Artevelde University College Ghent, Belgium

Abstract

Is it appropriate that the supervisor or tutor is also an evaluator? At Artevelde University College a significant effect on rating was found dependent on the role of the assessor (internal mentor, external mentor or independent assessor). Moreover a difference was found between a calculated score based on predetermined criteria (criteria based assessment) and a rating given by the assessor based on personal intuition (intuitive assessment). The results provide evidence for the separation between mentors or tutors and assessors when an evaluation of an undergraduate dissertation is required.

Introduction

One of the key questions in evaluating an undergraduate dissertation (thesis) is whether it is appropriate that the supervisor, mentor or tutor is also an evaluator or assessor. After an intensive process of coaching, it seems plausible that the supervisor has a certain relationship with the student that can disrupt his judgment when assessing a performance or an essay made by that student. At the Artevelde University College Ghent (field of education: midwifery) a study was conducted to determine the impact of mentoring on the assessment of a dissertation.

In recent literature, the dual role of being both supervisor (mentor) and assessor is generally not recommended. Teaching and assessing are fundamentally opposing roles. The dual role of teacher as supervisor on the one hand and assessor on the other hand, involves the risk that supervisors, to some extent, also assess themselves as joint decisions are made (Van Berkel, 2006). In the interaction between tutor and student, the supervisor may not be a reliable examiner (Wynants et al, 2011). Also students expect a disconnected assessment is more neutral because events during the educational process have no effect on the assessment (Elen et al, 2009). It seems that there is a greater assurance that the grant of the scores is more careful and more neutral in a decoupling of supervising and assessing (Elen et al, 2009).

Method

In this study the scores assigned by three different evaluators in the assessment of a bachelor thesis (undergraduate dissertation) were compared.

In the field of education of midwifery, the results of 57 graduate students were studied. The dissertation of each student was rated by 3 people; (1) An internal mentor or tutor who has guided the student and had frequent contact with the student, (2) An external mentor or supervisor from the field (often a hospital) who had regular contact with the student, (3) An external evaluator who has no contact with the student. Each assessor has the dissertation

report of the student. Each student also gives a presentation on his work. Both the dissertation and the presentation were assessed. At the end, a final score is formed.

Each assessor receives a score sheet. This form contains a number of criteria to be taken into account in the assessment. Each evaluator assesses the various criteria on a 4-point-Likert scale. Then, each evaluator may assign a score based on his intuition and experience. Afterwards, a standardized score was calculated based on the completed assessment form.

In the end, there were six main scores for each student to be compared (3 assessors with a score based on intuition and a calculated score). In addition, there were subscores for report, presentation, jury and process (the latter only judged by internal and external mentor). All data are analyzed and compared. A comparison of the frequency of the scores, a correlation study and a regression analysis were performed.

Findings

Results support the argument for a clear distinction between mentoring en assessing. There was a different scoring behavior between the internal mentor (internal tutor), the external mentor (external supervisor) and the independent assessor. Also, there was a difference between the calculated score based on predetermined criteria and the intuitive score given by the assessors.

First, the frequency of low scores (<10/20) and high scores (> 15/20) were compared. Intuitive scores were compared with the calculated scores.

Table 1: frequency of low and high scores (dissertation)

		Scores <10/20	Scores >15/20
Internal mentor	Intuition	0%	11%
	Calculation	14%	25%
External mentor	Intuition	4%	18%
	Calculation	12%	37%
Independent assessor	Intuition	14%	2%
	Calculation	46%	18%

Table 1 shows a clear effect: the more an assessor had contact with the student, the stronger the intuitive tendency to score no less than 10/20 (0% <4% <14%). No one of the 57 students received a score lower than 10/20 by his internal mentor. This indicates that internal mentors also assess themselves. The external mentor will rarely give a score below 10/20, but sometimes he or she does. The independent assessor frequently gives a lower score.

The calculated score is approximately equal to the internal and external mentor, so probably they assess the criteria approximately similar. The independent assessor is clearly more severe in judging the criteria. Possibly even too harsh, because 46% would score insufficient grants while intuitively only 14% will score insufficient grants.

Concerning the high scores (at least 16/20), internal mentors seem to be more cautious. Based on intuition, 11% of the student obtains a high score while 25% obtains a high score on the basis of the judged criteria (calculation).

External mentors give more frequently high scores when they are (intuitively) satisfied (18%). By calculation, even 37% of the students would get a high score. The independent assessor is very severe and careful when assigning high scores. On the basis of intuition the independent assessor would grant only 2% a high score. On the basis of the assessed criteria, this would be 18%.

Table2: Correlation between intuition and calculation

<u>Dissertation</u>	Report	Process
Internal mentor	0.70**	0.86**
External mentor	0.78**	0.80**
Independent Ass	0.63**	

** : correlation is significant at the 0.01 level

* : correlation is significant at the 0.05 level

Table 2 represents the correlations between the intuitive and the calculated dissertation score. These correlations are highly significant (for all assessors between 0.63 and 0.78). In the evaluation of the process, the relationship between intuition and calculation is even higher (0.80-0.86). All these high correlations indicate that there is certainly a relationship between the intuition of the assessor and the way he or she judges the criteria on the evaluation form. In comparison, table 3 indicates that the correlation between assessors is notable lower, although sometimes significant.

Table3: Correlation between assessor scores

<u>Intuition</u>	Internal mentor	External mentor	Independent ass
Internal mentor	1	0.39**	0.29*
External mentor	0.39**	1	0.22
Independent Ass	0.29*	0.22	1
<u>Calculation</u>			
Internal mentor	1	0.32*	0.18
External mentor	0.32*	1	0.24
Independent Ass	0.18	0.24	1

** : correlation is significant at the 0.01 level

* : correlation is significant at the 0.05 level

Table 3 indicates a strong correlation between the intuitive assessment of the internal and external mentor. But there is a remarkable difference when compared with the independent assessor. The independent assessor is the only assessor who does not know the student personally. In judging the criteria on the evaluation sheet (calculation), the difference is even more striking. There is a low correlation between the assessment of the independent assessor and the internal and external mentor.

Table 4: Frequency of scores: Process

		Scores <10/20	Scores >15/20
Internal mentor	Intuition	7%	9%
	Calculation	12%	46%
External mentor	Intuition	2%	19%
	Calculation	14%	40%

Table 4 indicates the frequencies of high and low scores by the internal and external mentor when assessing the learning process and cooperation process of the student. The external mentor intuitively grants less low scores and more high scores in comparison with the internal mentor. When the scores are calculated, we see the opposite: the internal mentor grants less low scores and more high scores.

Judging the process intuitively, the internal mentor is again more cautious and less extreme. Both low scores (7%) as high scores (9%) will be less present in comparison with calculated scores (on the basis of judging criteria). The external mentor is milder on the process and will grant less low scores and more high scores based on his intuition. When the judgment is based on clear criteria, the internal supervisor seems to be milder, or he remains more cautious because a low grant can be a sign of his own failure in counseling and guiding. On the other hand, a very high score can reflect too much self-praise as a tutor.

Table 5: Correlation internal mentor and external mentor: Process

<u>Process</u>	<u>Internal-external</u>
Intuition	0.36**
Calculation	0.26

** : correlation is significant at the 0.01 level

* : correlation is significant at the 0.05 level

Table 5 indicates that the correlation between the intuitive score of the internal mentor and external mentor is significantly high. Intuitively they have both the same opinion about the student and his or her process. However, the correlation between the calculated scores is low. When judging the students process, the internal and external mentor attach different value to different criteria or aspects. For example: for the internal mentor the learning process is very important. The external mentor, however, gives more value to the cooperation process. This indicates that the internal and external mentor evaluate in their own way the process of the student. Even though their overall impression is generally similar, they pay attention to other aspects and attach different values to these aspects.

Table 6: Frequency of scores: Presentation

		Scores <10/20	Scores >15/20
Jury	Intuition	4%	2%
	Calculation	5%	30%

Table 6 represents the results of the assessment of the presentation. A common score is given by the assessors as a jury. It is noteworthy that intuitively few low or high scores are given. Based on calculation, however, 30% would get a high score compared with 2% intuitively. This results show that the judgment of a presentation is much more guided by an overall feeling or impression than by predetermined criteria.

A regression analysis was performed to determine which part in the evaluation process is decisive for the final grade. If a regression analysis is performed on the calculated scores, the final grade for a dissertation can be well predicted ($r^2 = 88\%$, explained variance) by 3 factor scores. The 3 most predictive scores are: (1) the dissertation assessment by the independent reader, (2) the assessment of the process by the internal mentor and (3) the dissertation evaluation by the external mentor. This remarkable result advocates that an independent

assessor (who can be seen as more neutral) and an external mentor (who guides the student less intensively) can better judge the evaluation of the dissertation report than the internal mentor. The internal mentor can be helpful by judging the (learning) process of a student. A dissertation presentation by the student provides little additional information or value when scores are calculated on the basis of predefined criteria.

A regression analysis on the intuitive scores showed 4 factors that can well predict the final result ($r^2=0.96$). In this case, the most predictive score is the assessment of the presentation. Second: the evaluation of the process by the internal mentor. Third: the judgment of the dissertation by the independent assessor. And fourth: the assessment of the dissertation by the external mentor. It is clear that an intuitive judgment is highly influenced by the impression that a student leaves, both during the presentation and during the process. On the other hand, it is also possible that the impression of the dissertation has an influence on the review of the presentation by the student.

Discussion

The results indicate that the more someone is familiar with a student and the stronger the supporting role, the lower the intention to give a low score. Low scores are less common when based on intuition than when calculated on the basis of predefined criteria. Also high scores are less common based on intuition. Intuitive judgments seem to be more cautious and less extreme than calculated scores. Calculated scores show a clear difference based on the familiarity with the student: the more familiar, the less strict is the judgment of the criteria.

Different assessors look at different ways to the performance of a student. Familiarity and contact with the student seem to have an important influence. When intuitively assessing, the overall impression will play an important role. Different assessors attach different value to other aspects. If one intends to make a judgment on predefined criteria, an independent assessor seems the most neutral judge (objectiveness). Although, they sometimes are more severe than they would judge intuitively.

Objectiveness is not equal to reliability (Van Hout, 2006). Personal influences in a review can never be ruled out completely. Different evaluators can catch or reduce these influences. Reaching a consensus with several assessors increases the reliability of the assessment. A process of moderation and reconciliation is needed to increase reliability when different assessors are involved. The assessor will always play a determining role in the evaluation. How concrete and unambiguous criteria are formulated, there is always an appeal to the interpretive power of the assessor. A mentor can only be a reliably assessor if he sufficiently can step away from the events during the process at the time of the evaluation. (Wynants, 2011).

The tension between supervision and evaluation remains a difficult issue because it also has benefits when an evaluator is also a mentor. It also can be discussed if the personal progress of a student should or should not be taken into account when assessing his achievements. However, the results of this study provide evidence for the separation between supervisors or mentors and assessors when neutral evaluation of an undergraduate dissertation is required.

Practical implications

These findings have led to a revision of the evaluation process of the undergraduate dissertation in different training programs at the Artevelde University College. A digital assessment tool called DiSco (Digital Scoring) was developed. When constructing this tool, findings from this study were taken into account. In DiSco, fully detailed criteria can be saved and be assessed on a 4-point-Likert scale. A variety of evaluator roles can be provided. Every criterion can be linked to a certain role. For example: an internal mentor judges only criteria related to the learning process and an independent assessor only judges criteria linked on the dissertation report. Assessors should not give (intuitive) scores any more. These scores are automatically calculated by the tool. Different criteria can have different weights. Every criterion has the same value for every assessed student and the fixed calculation ensures that the score for each student is achieved in a similar way. Sufficient thinking about “who will assess what” and determining the value of each aspect or criterion seems to be necessary.

References

- Bray L, Nettleton P. (2007) Assessor or mentor? Role confusion in professional education. *Nurse Educ Today*. 2007 Nov;27(8):848-55. Epub 2007 Feb 12.
- Carifio, M. S., & Hess, A. K. (1987). Who is the ideal supervisor? *Professional Psychology*, 18, 244-250.
- Elen, J., Sasanguie, D., Coens, J., Clarebout, G., Van den Noortgate, W., Vandenabeele, J., de Fraine, B. (2009). Ontkoppelen van begeleiden en summatief beoordelen in het hoger onderwijs: een aanzet tot discussie. *Tijdschrift voor hoger Onderwijs*, 27(3),157-170.
- Van Berkel, H.,& Bax, A. (2006). *Toetsen in het hoger onderwijs*. Houten: Bohn Stafleu Van loghum.
- Van Hout, H.(2006), *Vernieuwing in het hoger onderwijs*. Assen: Van Gorcum.
- Wijnants, L., Hermans,I., & Vandeput, L.(2011). *Innovatief evalueren in het hoger onderwijs: Leerwegonafhankelijk summatief toetsen*. Lannoo Campus.