#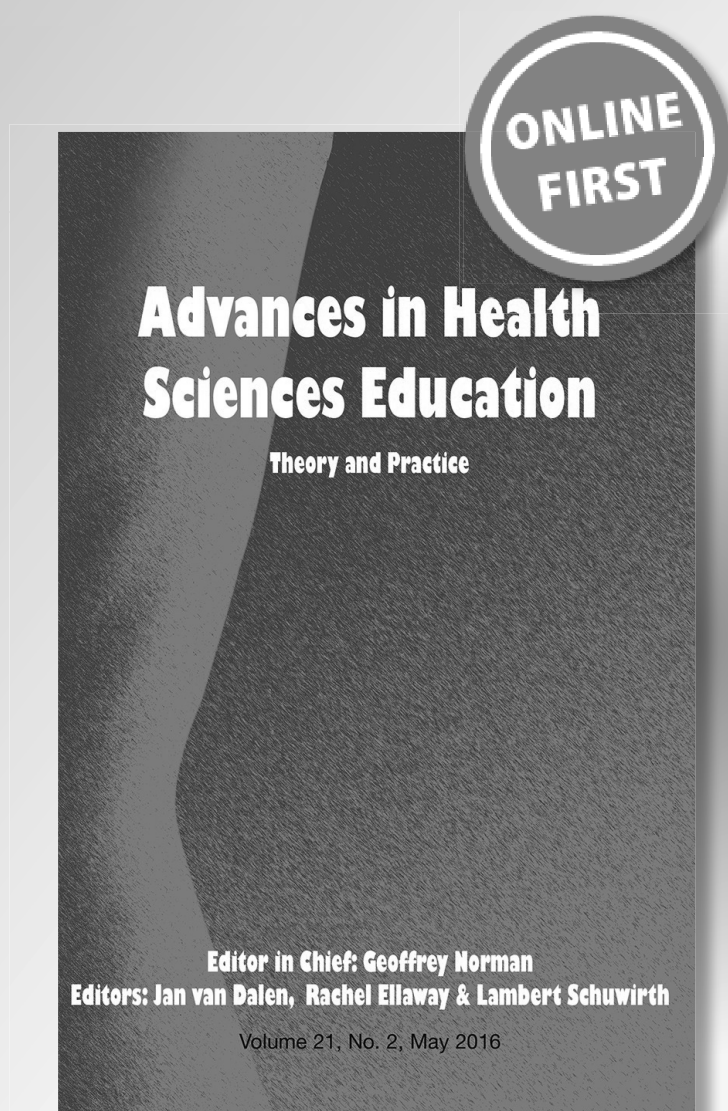 The validity of student tutors' judgments in early detection of struggling in medical school. A prospective cohort study

Lotte Dyhrberg O'Neill, Anne Mette Morcke & Berit Eika

Advances in Health Sciences Education

Theory and Practice

Editor in Chief: Geoffrey Norman
Editors: Jan van Dalen, Rachel Ellaway & Lambert Schuwirth

Volume 21, No. 2, May 2016

Springer

Springer

CrossMark

The validity of student tutors' judgments in early detection of struggling in medical school. A prospective cohort study

Lotte Dyhrberg O'Neill[1] · Anne Mette Morcke[1] · Berit Eika[2]

Abstract Early identification and support of strugglers in medical education is generally recommended in the research literature, though very little evidence of the diagnostic qualities of early teacher judgments in medical education currently exists. The aim of this study was to examine the validity of early diagnosis of struggling in medical school based on informal teacher judgements of in-class behavior. The study design was a prospective cohort study and the outcomes/truth criteria were anatomy failure and medical school drop out. Six weeks into an anatomy course, student tutors attempted to identify medical students, who they reckoned would fail the anatomy course or drop out, based on their everyday experiences with students in a large group educational setting. In addition, they were asked to describe the indicators of struggling they observed. Sixteen student tutors evaluated 429 medical students for signs of struggling. By week six, the student tutors were able to detect approximately 1/4–1/3 of the students who eventually failed or dropped out, and for ¾ of the strugglers they identified, they were correct in their judgments. Informal student tutor's judgements showed incremental validity for both outcomes when controlling for grades obtained in preceeding exams. Lack of participation, lack of commitment, poor academic performance, poor social interactions and general signs of distress were the main indicators of struggling identified. Teachers' informal judgements of in-class behavior may be an untapped source of information in the early identification of struggling medical students with added value above and beyond formal testing.

Keywords Attrition · Drop out · Education · Failure · Medical Student · Performance Assessment · Prediction · Undergraduate

✉ Lotte Dyhrberg O'Neill
lotte@cesu.au.dk

1 Centre for Health Sciences Education, INCUBA Science Park Skejby, Aarhus University, Palle Juul-Jensens Boulevard 82, building B, 8200 Århus N, Denmark

2 Rector's Office, Aarhus University, Aarhus, Denmark

⧆ Springer

Introduction

Medical students and medical doctors who fail to progress as planned in their educational programs have been given many names, such as: strugglers, problem learners, at-risk students, problem residents, troublesome learners, disruptive students, residents in difficulty, impaired physicians etc. (Yates and James 2006; Tabby et al. 2011; Steinert 2013; Winston et al. 2014). In this paper we will use the term "struggler" to mean a student who progresses with difficulty, which may or may not involve a strenuous effort. The prevalence of strugglers is not often reported, though a few studies have suggested struggler rates of 3–10 % in post-graduate residency programs (Tabby et al. 2011; Steinert 2013; Zbieranowski et al. 2013), and 10–15 % in undergraduate medical education (Yates and James 2006; Yates 2011). Struggling in medical school is probably strongly associated with attrition, even though surprisingly few good quality studies have actually documented this link (Hojat et al. 1996; Stetto et al. 2004; Yates and James 2006; O'Neill et al. 2011; Yates 2011; Mørcke et al. 2012; Fortin et al. 2015). Attrition constitutes wasted efforts for the student in question, the educational institution, and society—in addition to the negative effects it may have on self-efficacy and confidence. There is also some evidence that there may be a price to pay for patients associated with unresolved struggling in medical school, which also makes a strong case for early identification and follow up of struggling medical students (Papadakis et al. 2004, 2005; Teherani et al. 2005; Brenner et al. 2010; Yates and James 2010). It has been shown that US physicians who have been the subject of disciplinary proceedings are more likely to have struggled at medical school, with this struggling taking the form of repeated episodes of irresponsibility and diminished capacity for self-improvement (unprofessional behavior), as well as poor academic performance (Papadakis et al. 2004, 2005; Teherani et al. 2005). Even the mildest type of professional disciplinary cases may have the potential to adversely affect patients (Papadakis et al. 2005). In the UK, researchers have also found that doctors sanctioned for professional misconduct had increased risks of having failed exams early in the preclinical years of medical school (Yates and James 2010). Clearly, strugglers represent both a practical challenge and an ethical dilemma for both teachers and program directors (Yao and Wright 2001; Roberts et al. 2012; Zbieranowski et al. 2013). Roberts et al. (2012) for example, highlighted the challenges and dilemmas involved in balancing the needs of the individual struggler with the needs of the healthcare system and the people within it. And similar balancing acts between the needs of strugglers and non-strugglers are necessary in non-clinical educational settings. Apart from such resource-related challenges and dilemmas, a number of barriers may exist which prevent teachers from approaching strugglers. Teachers may feel that it is 'not their role', that they lack the sufficient skills to do so effectively, that they could be opening a can of worms which could potentially make things worse (Evans et al. 2010; Steinert 2013), and that the resultant emotional distress and self-doubt incurred on the struggling student could be inimical to learning (Mazor et al. 2005). While these reservations might look like excuses or seem irresponsible to some, they also represent valid and important doubts on behalf of teachers. Even so, over the years many researchers have stressed the critical importance of early identification and early support of strugglers (Yao and Wright 2001; Evans et al. 2010; Yates and James 2010; Steinert 2013; Winston et al. 2014). Early identification and support has been presented as 'an important investment in the development and training of future health professionals' (Steinert 2013), and as the 'gold standard for educational supervision' (Evans et al. 2010). However, it is

important to consider the evidence to support the notion that teachers can accurately identify strugglers before official failure is manifest and damage is done.

Teacher predictions of failure

Teacher judgments have generally been found to be correlated to subsequent student performance in pre-school and primary education (Wijnia et al. 2013). However, there seems to be very few studies on teachers' ability to predict in advance student success or failure in higher education (Harackiewicz et al. 2002; Van den Berg and Hofman 2005; Georg 2009; Wijnia et al. 2013). We found six existing studies reporting on teacher judgments' validity in higher education—all appear to have taken place in problem-based learning (PBL) settings (Kaufman and Hansell 1997; Whitfield and Xie 2002; Van de Watering and Claessens 2003; Loyens et al. 2007; Adam et al. 2012; Wijnia et al. 2013). Three studies seem to indicate that PBL tutors in higher education were able to predict students' academic achievement and dropout (Van de Watering and Claessens 2003; Loyens et al. 2007; Wijnia et al. 2013). The three other studies examined PBL facilitators' ability to predict written exam scores in medical education, and found that PBL facilitators' predictions of grades were only weakly to moderately correlated to the value of the subsequent exam scores (Kaufman and Hansell 1997; Whitfield and Xie 2002; Adam et al. 2012). While the ability to judge academic performance is obviously important in detecting struggling early, we would question whether teachers need to be able to predict the specific value of future exam grades for all students in advance. It may be more relevant to be able to diagnose broader performance categories such as examination failure and drop out. We found no studies evaluating the diagnostic quality of medical teachers' informal judgments of strugglers very early in a medical program.

In summary

It is recommended that we identify or diagnose strugglers early in medical education, but currently very little evidence of the diagnostic qualities of early tutor judgments in medical education exist, nor does there seem to be any studies of teacher judgements in non-PBL settings. The aim of this study was therefore to examine the validity of early diagnosis of struggling in a medical school. The objectives were (1) to examine the validity of an early diagnosis of exam failure and dropout in a non-PBL medical curriculum based on teachers' informal judgments, and (2) to examine which observed indicators of struggeling the teachers based their judgments on.

Method

This study was a prospective cohort study examining the validity of teacher judgments as an early diagnostic tool of struggling medical students.

Participants

All tutors teaching the course Anatomy 1 at Aarhus University (AU) medical school in 2012 were invited to participate in the survey 'Early Identification of Strugglers'. Anatomy 1 was scheduled in the first semester of the medical program. The objects of the tutor

judgments were all medical students who had registered to participate in this anatomy course. The students had some of the highest (top 5 %) pre-university grade point averages in the country. In addition, their educational fees were fully publically funded, they were all entitled to student grants of around $ 10,800 per year, and additional student loans of ca. $ 5500 per year to cover living expenses. The anatomy tutors participating in this study were also medical students. The requirement for being hired as an anatomy tutor was having completed the same anatomy course with outstanding grades at any time in the past. The anatomy tutors are not involved in the final examination of students in the course, which is administered by the senior lecturers (anatomists) responsible for the course.

Setting and exposure

Each year AU medical school admits around 450–500 new students which are divided into either September or February starters; therefore this anatomy course which is scheduled in the first semester of the medical curriculum is delivered twice a year. The course lasts 15 weeks and consists of a mixture of lectures by senior anatomists, and tutorials and dissection classes run by the medical student tutors. The anatomy tutors invited to participate in this study each taught their own class of approximately 25–30 students for 32 tutorial hours (2 h/day times 16) and 12 dissection hours (3 h/day times 4), i.e. approximately 44 h in total during a 6 week period before attempting to identify strugglers for this study. This corresponded to a point in time which was approximately half-way through the course. The topics covered during this period of time were basic introductory anatomy and musculoskeletal anatomy of the upper extremity in particular. The delivery of teaching in these tutorials is probably best described as interactive lectures in a large group setting supplemented with practical small-group exercises in the dissection lab. Following the course, students are assessed on their anatomy knowledge with a 1 h written test consisting of both a spot test element and a short answer test element, the latter consisting of 44–64 individual questions. Students' anonymity was secured in the marking process, and all students were double marked by both an internal and an external examiner.

Instrument

We introduced the anatomy tutors to the study at a teacher meeting scheduled before semester start in the early spring of 2012. Our aim was to hear if they thought it possible to identify strugglers at all, to gauge their willingness to participate, to hear and discuss any ethical issues that would surface and to discuss a suitable time of survey. Tutors were generally positive on all aspects discussed and proposed. Therefore, after 6 weeks of teaching, each tutor was sent a link to an electronic survey asking them to evaluate any strugglers in their tutorial group. On the first page of the survey the aim of the study was explained again, as were the anonymity levels of the analyses and reporting of results. On the following page a struggler was defined as a student who was in danger of either failing the following anatomy exam or dropping out of the program, and the tutors were asked if there were any strugglers in their group. If they said yes, they were asked to choose the person in question on a drop-down list. They were then asked to answer yes or no to the following question: "As far as I can see, this person is in danger of not passing the first following exam in the Anatomy 1 course". If the answer was yes, the respondent was asked: "Why do you think, this person is in danger of not passing the first following exam in the Anatomy 1 course? (Describe with free text or key words)." The subsequent question was a yes/no question and read: "As far as I can see, this person is in danger of

dropping out of the program at some point in time". If the tutor answered yes to this question, they were again asked to explain in free text or with key words why they thought so. They were then asked if there were more strugglers in their group, and if the answer was yes, the whole process would be repeated until they answered no to identifying any more strugglers, at which point the questionnaire would terminate. We used the free open source software LimeSurvey as our survey tool, which stored data on our own servers.

Predictors of performance

We included four predictors of performance in modeling the medical school performance outcomes examined (anatomy failure and drop out), they were: tutor judgments, pre-university grade point averages (pu-GPA), the type of pre-university exam and Anatomy grades.

Tutor judgments

Based on the tutor responses to the survey, we categorized students as either tutor judged anatomy failures or not, and tutor judged dropouts or not.

Pu-GPA

Pu-GPA was the average grades obtained in any type of upper secondary education exam prior to university admission as measured on the Danish 7-point grade scale, which consists of the following values and performance categories: 12 (excellent), 10 (very good), 7 (good), 4 (fair), 2 (minimally adequate), 00 (inadequate) and −3 (unacceptable), equivalent to the marks A, B, C, D, E, F_x and F on the European Credit Transfer System (ECTS) scale.

Pre-university exam types

The pre-university exam types were categorized into either the Danish 'Gymnasium' exam, 'other' exam types (the higher preparatory examination, the higher commercial examination, the higher technical examination, or the International Baccalaureates etc.) or as 'missing' data. These exams are typically taken by 16–19-year-old students and are all at the upper secondary educational level. The common objective of these programmes is to prepare young people for higher education.

Anatomy grades

The course grades given in anatomy were on the same scale as the pre-university grades. A grade below 2 (minimally adequate) meant failing the exam.

Outcomes/truth criteria

We extracted data on anatomy grades and program status from the Student Administrative System (STADS), in order to form the two outcome variables of struggling: 'anatomy failure' and 'dropout', which were our 'truth criteria'.

Anatomy failure

Failing anatomy was defined as not obtaining the minimum pass grade in the first exam administered following the course, or not attempting that exam at all (i.e. postponing the exam for any reason).

Dropout

In the STADS database students are registered with dates for start, and termination or completion of studies. Drop out was defined as having terminated studies at AU by the end of March 2015 for any reason (withdrawal, dismissal, or transferal). Non-dropouts were students who were still active (delayed or on time). Hence, the follow-up time was 3 years for the spring cohort and 2½ years for the autumn cohort. We know from previous research, that we may typically expect more than 80 % of our dropout to occur within 2 years from study start (Mørcke et al. 2012).

Analysis

The quality of the diagnostic tool 'teacher judgments' of struggling

For each of the two outcomes ('anatomy failure' and 'dropout') a 2×2 tabulation of tutor judgments versus the outcome was constructed. To evaluate the validity of the diagnostic tool of tutor judgments of struggling, we calculated indices to describe the diagnostic quality, of particular interest relative to our research question were: the sensitivity and the positive predictive value. The sensitivity was calculated as the number of: true positives/(true positives + false negatives) or $TP/(TP + FN)$. The positive predictive value was calculated as the number of: true positives/(true positives + false positives) or $TP/(TP + FP)$ (Altman 1991; Juul 2004).

Incremental validity

In addition we evaluated the incremental validity of tutor judgments by testing models of tutor judgments as predictor variables of struggling while controlling for examination results. Since our outcomes were bivariate (anatomy failure/non-failure and dropout/non-dropout) univariate and multivariate logistic regression was used. All predictor variables were screened for collinearity and zero cells by inspecting matrix plots, box plot and 2×2 tabulations, and by calculation of a tolerance statistic for collinearity between predictors (Menard 2002). We examined multivariate models using backwards hierarchical elimination (Kleinbaum and Klein 2002). Descriptive summary statistics and analyses were performed using STATA/IC 12.

Qualitative analysis

The open-ended questions of the survey which invited respondents to describe the indicators of struggle observed were analyzed qualitatively with content analysis. We took an explorative approach which meant using inductive coding. Firstly, one researcher (LON) coded all themes with key words and subsequently grouped these into meaningful

categories. The coding was then validated using researcher triangulation with a second researcher (AMM). The final categories were agreed by all authors.

Ethics

This project is exempt from ethics review by the regional (biomedical) ethics committee as surveys, database studies and quality assurance studies do not require their permission. We therefore obtained permission to conduct the research study from Aarhus University Faculty of Health Sciences. For research projects like this, which involve university students, their names, their university e-mail addresses, their student id numbers, their performances at university (i.e. grades and progressional status etc.) permission from the Danish Data Protection Agency is not required according to section 48 in the current Personal Data Act no. 429 of 31 May 2000, because these particular types of data are not considered 'sensitive data' as defined in subsections 7 and 8. Examples of what *is* considered 'sensitive data' are data on: race or ethnicity, political or religious or philosophical convictions, union membership, health or sexuality. However, research projects like the study presented here must still adhere to the ordinary rules in the Personal Data Act regarding appropriate data handling in research projects, i.e.: data must be stored safely, they must not be handed over to third parties without permission, they must not be used for other purposes than the intended research, data should be deleted or anonymized after completion of the project, and results should be published in a form which does not allow identification of individuals. In this research project we adhered to these requirements.

Results

In 2012, a total of 481 medical students (237 in the spring semester and 244 in the autumn semester) embarked on the Anatomy 1 course at AU medical school. All 481 students were assigned to one of 18 large groups. We received responses to our survey from 16 of the 18 group tutors who evaluated 429 of the students all together resulting in a response rate of 89 %. A total of 37 strugglers were identified by the tutors corresponding to 9 % (37/429) of the population evaluated. Of the 37 strugglers identified, 22 were thought to be in danger of both failing anatomy and dropping out, 10 were perceived to be in danger of anatomy failure only, whereas 5 were predicted to drop out without failing anatomy first. In the end, 23 % (99/429) of the students failed their first attempt at Anatomy 1 (Table 1), and 14 % (59/429) of them had dropped out of medical school by March 2015 (Table 2).

Table 1 Early student tutors' judgments of anatomy failure

Anatomy failure			
Tutor judgment	Fail	Pass	Total
Fail	24 (TP)	8 (FP)	32
Pass	75 (FN)	322 (TN)	397
Total	99	330	429

Sensitivity $= TP/(TP + FN) = 24/(24 + 75) = 0.24$, *specificity* $= TN/(FP + TN) = 322/(8 + 322) = 0.98$, *positive predictive value* $= TP/(TP + FP) = 24/(24 + 8) = 0.75$, *negative predictive value* $= TN/(FN + TN) = 322/(75 + 322) = 0.81$

Early teacher judgements as a diagnostic tool

The sensitivity of tutors informal judgements of subsequent anatomy examination failure was 0.24 meaning that the probability of having been diagnosed as a struggler by week 6 was 24 % amongst the students who ended up failing anatomy, in other words: the majority (ca. ¾) of the failing students were not diagnosed as strugglers by tutors at this early point in time. The positive predictive value was 0.75, which means that the probability of failing the anatomy exam for students judged to be strugglers by tutors by week 6 was 75 % (Table 1). The sensitivity of early tutor judgements of dropout was 0.34, while the positive predictive value was 0.74 (Table 2).

Incremental validity

Table 3 describes the student sample on selected variables. As a single (univariate) predictor, tutor judgements at week 6 seemed to explain more variance (9.2 %) than pre-university grades did (4.1 %) for the outcome of anatomy failure (R^2's in Table 4). In the final model in Table 4, tutor predictions survived as a significant independent predictor of anatomy failure, and it added substantial incremental validity ($R^2 = 0.119–0.041 = 0.078$) compared to a univariate model containing only pre-university grades as a predictor of anatomy failure. Even when controlling for pre-university grades (general academic ability), the odds of failing anatomy were almost 12 times higher for students identified as strugglers by their tutor at week 6 compared to the students who were judged non-strugglers (Table 4).

At week 6 of the anatomy course, only tutor predictions and pre-university grades survived as significant predictors of subsequent drop out, and tutor judgements added substantial incremental validity ($R^2 = 0.171–0.028 = 0.143$) to the univariate model containing only pre-university grades (model 1, Table 5). For strugglers identified as being at risk of dropping out at week 6 by their tutor, the odds of subsequent drop out were 24 times that of the other students when controlling for differences in general/prior academic ability. After the anatomy exam only the obtained anatomy exam grades and tutor judgements survived as significant predictors of drop out, and tutor judgements added incremental validity ($R^2 = 0.249–0.224 = 0.025$) to a model with only anatomy grades as a univariate predictor. We found no evidence of collinearity between predictors.

Reported indicators of struggling

We analyzed tutors descriptions of signs of struggle, and identified 5 categories, which together covered all meaningful descriptions given for both outcomes (anatomy failure and

Table 2 Early student tutors' judgments of dropout

Drop out			
Tutor judgment	Dropped out	Persisted	Total
Dropout	20 (TP)	7 (FP)	27
Non-dropout	39 (FN)	363 (TN)	402
Total	59	370	429

Sensitivity = TP/(TP + FN) = 20/(20 + 39) = 0.34, specificity = TN/(FP + TN) = 363/(7 + 363) = 0.98, positive predictive value = TP/(TP + FP) = 20/(20 + 7) = 0.74, negative predictive value = TN/(FN + TN) = 363/(39 + 363) = 0.90

Table 3 Medical student sample studied (n = 429)

	N
Gender	429
Males	147
Females	282
Pu-GPA	399
	Mean score 10.73
	Score SD 1.19
Pu-exam type	429
Danish Gymnasium	324
Other	75
Missing	30
Tutor judged anatomy failures	429
Failures	32
Non-failures	397
Tutor judged dropouts	429
Dropouts	27
Non-dropouts	402
Anatomy grades	396
	Mean score 5.03
	Score SD 3.45
Anatomy failures	429
Passed	330
Failed	99
Dropouts	429
Dropped out	59
Persisted	370

Table 4 Incremental validity of early student tutors' judgements of anatomy failure

	Univariate analyses			Final model	
	OR [CI 95 %]	p	Pseudo R^2	OR [CI 95 %]	p
Tutor judged anatomy failure	12.88 [5.56–29.82]	0.000	0.092	11.71 [4.80–28.69]	0.000
Pu-GPA	0.66 [0.54–0.81]	0.000	0.041	0.72 [0.59–0.88]	0.001
Pu-exam type		0.016[a]	0.017	–	
Other	2.20 [1.27–3.79]	0.005			
Missing	0.98 [0.38–2.49]	0.962			
N_{obs}				399	
Pseudo R^2				0.119	

The Danish Gymnasium exam is the reference category for pu-exam type

OR odds ratio, *CI 95 %* 95 % confidence interval of OR, *p p* value

[a] The *p* value for the overall variable (pu-exam type)

Table 5 Incremental validity of early student tutors' judgement of medical school dropout

	Univariate analyses			Model 1 (At week 6)		Model 2 (After the anatomy exam)	
	OR [CI 95 %]	p	Pseudo R^2	OR [CI 95 %]	p	OR [CI 95 %]	p
Tutor judged dropouts	26.59 [10.57–66.93]	0.000	0.165	24.43 [9.31–64.15]	0.000	4.57 [1.50–13.96]	0.008
Pu-GPA	0.72 [0.58–0.99]	0.002	0.028	0.79 [0.63–0.99]	0.042	–	
Pu-exam type		0.005[a]	0.034	–		–	
Other	2.48 [1.33–4.61]	0.004					
Missing	0.25 [0.03–1.91]	0.182					
Anatomy grade	0.64 [0.52–0.78]	0.000	0.224	–		0.68 [0.56–0.83]	0.000
N_{obs}				399		396	
Pseudo R^2				0.171		0.249	

The Danish Gymnasium exam is the reference category for pu-exam type

OR odds ratio, *CI 95 %* 95 % confidence interval of OR, *p* p value

[a] The *p* value for the overall variable (pu-exam type)

drop out). Tutors noticed: problems with *participation* in the course, problems with *commitment*, *academic problems*, *social interaction problems*, and signs of *distress* (Table 6). Overall, problems with participation and commitment were the most common signs of struggling observed followed by academic problems (Table 6).

Discussion

After 6 weeks of teaching, the anatomy tutors were able to detect approximately 1/4–1/3 of the struggling students, and for ¾ of the strugglers they identified, they were correct in their perceptions. Informal tutor judgements showed incremental validity when controlling for grades obtained in previous exams. Lack of participation, lack of commitment and poor academic performance were the most common indicators of struggling identified.

Tutors in this study were for the most part justly alarmed, when they were alarmed (high positive predictive values), even though they tended to under-detect the strugglers (low sensitivities). A low sensitivity of tutor observations for strugglers is unfortunate, but acceptable because we are not dealing with an expensive screening procedure with added deleterious side effects. In the situation, we are merely dealing with observations tutors do privately in their everyday interactions with students anyway. One could therefore argue that there was nothing to lose but potentially something to be gained. As our focus was whether tutors were justified in approaching perceived strugglers with the intent to suggest support, the high positive predictive value is more important than the low sensitivity. For comparison, we calculated the sensitivity for tutor predictions of dropout based on published data. Wijnia et al. (2013) asked 15 PBL tutors to judge 211 psychology students on their individual likelihood on a 0–100 % scale of completing the bachelor programme. Their data reveals a sensitivity of 0.44 for tutor judgements of future drop out after 5 weeks of PBL, and a corresponding positive predictive value of 0.89, both of which are somewhat higher, but still comparable to the values we found. However, only the subset of data on the students (n = 81) with the extreme ratings (i.e. in the upper or the lower

Table 6 Observed signs of struggling: categories, descriptions and number of comments

Categories	Examples of descriptions	Comments for predicted anatomy failures (n)	Comments for predicted dropouts (n)	Total (n)
Participation problems	Passive, lack of initiative, have to drag answers out of the student, lack of active participation, mentally absent, does not attend all classes, very irregular attendance, absent from most tutorials, stopped attending tutorials, does not attend, never turned up	17	8	25
Commitment problems	Not quite the same interest, not so dedicated, lack of motivation, lack of interest, wants to try the sports science program instead, says he/she is more interested in the psychology program, has not been able to afford books, great distance between his/her house and the institution, lack of diligence, lack of effort, lack of preparation, appears unprepared, finds it very difficult to pull herself together and prepare, finds it difficult to accept the great content load and work load, already dropped out	13	10	23
Academic problems	Difficulties with learning, lack of study skills, finds the subject enormous and confusing, and cannot prioritize important content over less important content, lack of overview, exaggerated attention to minor details, often gives wrong answers, and it seems like she is unable to understand the questions, can rarely answer even simple questions, gives completely meaningless answers to questions, lack of knowledge, academic difficulties, academically inferior, academically challenged	16	4	20
Social interaction problems	Not that well integrated in the group, not socially well integrated, less socially attached, little social contact, very quiet, asocial, has announced that she wants to prepare on her own, cannot find the words, language barrier, language problems, difficulties with the Danish language, immature, too much focus on social life	5	7	12
Signs of distress	Does not believe in own proficiency, very insecure, has made several prior attempts and seems very stressed out, great insecurity, which may be a burden in the long run	5	2	7

The numbers in the table is the number of comments in the categories for each of the two outcomes

quartiles on the scale) were reported. It is likely, that the outcomes of the students receiving the extreme tutor ratings were easier to detect and predict precisely, and that this could have inflated the values somewhat.

In our study the tutors were near-peers and not expert teachers. However, we think both groups may have potential advantages when it comes to making early informal judgments about their students. Near-peers' advantages could be that they experience more honest behaviours and responses from their students, when they are not in a position of power. It is also possible that near-peers' communication with and understanding of their students is superior because of a larger degree of social and cognitive congruences (Yu et al. 2011). Both conditions could give rise to better insights into students' situations which could improve their predictions. On the other hand, seasoned teachers may have superior academic and teaching skills, which may make the teaching task less of a challenge. This could permit increased attention to student behaviours and reactions in class. In addition, experienced teachers may have years of practice in spotting struggling students, and opportunities to learn from having their informal judgments checked in the exams. Such circumstances could give them an edge in the identification process compared to novice tutors. At the present, there is simply not enough evidence in the literature to suggest which type of tutor (near-peer or expert) makes the more precise early predictions of failure

Another likely influence on teachers' ability to identify struggling students may be the student group sizes and the educational activities involved. Wijnia et al. (2013) examined teacher predictions in PBL groups with maximum 15 students in which the teacher (with a degree in psychology) had a facilitatory role. In such a setting, there may be more time to observe the performance of each individual repeatedly and in greater depth. In contrast, our tutors were students themselves teaching larger groups of 20–30 students with a much more traditional, teacher-centered approach. This may have resulted in fewer opportunities for our tutors to get to know all students in their group well, which could also explain the somewhat lower values in our setting. Loyens et al. (2007) also examined PBL tutor ratings of Dutch psychology students' observed learning activities (preparation, participation, and role as chair and scribe) and found they were very strong predictors of academic achievement and a good predictor of dropout in a structural equation model.

The timing of the identification of struggling students may also be a likely influence on sensitivity and prediction. A recent study reported on the administration of an exam already in week 2 of the program in order to identify at-risk students in a US-Caribbean medical school (Winston et al. 2014). Of the students who failed this early exam, 65 % went on to fail one of the first three semesters in the program, corresponding to a positive predictive validity of 0.65. This is slightly lower than what we found, which could perhaps be explained by the very early test administration influencing the content validity of the test negatively. It is also possible, that formal testing does not capture other important factors for success, such as the levels of participation, commitment and social integration.

Incremental validity

The non-expert tutors in this study were able to make predictions based on informal observations at an early stage of this non-PBL program in a manner which was independent of and added to the information available from examination results (Tables 4 and 5). This is evidence of the importance of not just using formal test results and administrative data for identification of at-risk students. It shows that tutors' informal everyday subjective observations has added value and is worthy of some trust. In Danish universities struggling students have traditionally been handled by administrators and counsellors in the

Table 7 Recent examples of reported signs of struggling or dropout in medical school

Author	Type of paper	Outcome	Signs of struggling presented
Maher et al. (2013)	A retrospective cohort study. N = 779	Factors associated with attrition in medical school	Absenteeism Leave of absence Social isolation Academic struggling Depression/Psychological morbidity
Mørcke et al. (2012)	A retrospective cohort study of medical students at Aarhus University. N = 639	Factors associated with attrition in medical school	Low grades in medical school Leave of absence
Roberts et al. (2012)	Expert discussions of a composite case derived from interviews of program directors and review of resident records in the UK	Residents with performance problems	Repeated interpersonal problems (rude, impatient, condescending, arrogant, and overconfident behaviours towards peers, staff and patients) Irresponsible behaviour (being late, absenteeism, lack of follow up, blaming others) Non-collaborative behaviour (dumping chores on others, overruling peers) Failure to improve (repeated reprimands ineffective
Hayes et al. (2011)	UK/international consensus groups profiling support seeking medical students	Medical students seeking support	Poor learning skills (often strong past academic performance) Poor organisational skills (e.g. problems with performance, attendance and deadlines) Poor mental health Immaturity Poor insight (poor professional or academic performance, strong self-belief, external locus of control/blaming of others, impervious to feedback, persistent behaviour patterns) Major personal crisis
Tabby et al. (2011)	A survey of 126 neurology residency program directors in the US	Problem residents in neurology	Inappropriate interaction Poor clinical judgment Late/absent Poor patient management Problems with humanism, moral, emotions, language and illegal behaviour Bad attitude Untrustworthy Slow
Yates (2011)	A retrospective cohort study at University of Nottingham Medical school (N = 1188)	Medical students at risk of failure to thrive	Failure of 3 or more examinations per year An overall average score of <50 % Health difficulties Social difficulties Failure to complete Hepatitis B vaccination on time Remarks about poor attitude or behaviour A lack of professional behaviour

Table 7 continued

Author	Type of paper	Outcome	Signs of struggling presented
Evans et al. (2010)	A discussion paper by experienced support faculty on the importance of early identification and intervention	Students and trainees in difficulty	Poor attendance Poor clinical skills, inconsistent with stage of training Poor knowledge, inconsistent with stage of training Lack of or poor relationships with patients The student who is the 'ghost at the back' (present in body only) Lack of collaborations with peers Worrying non-verbal communication Behaviour consistent with disenfranchisement from professional identity Anxiety

university support system, in which students seek support on their own initiative. The problem is that students with the most severe shortcomings may be the ones with the least amount of self-insight, coping abilities, and help-seeking behaviour (Dunning et al. 2003; Cleland et al. 2005; Devoe et al. 2007; Winston et al. 2010). It is wellknown from research in both psychology and medical education, that it is particularly difficult for the poorest performing students to recognize their own incompetence in advance (Hodges et al. 2001; Dunning et al. 2003; Cleland et al. 2005). Feedback and support from others are generally considered very important for learning in medical education by both educationalists and strugglers (Challis 2000; Cleland et al. 2005; Kilminster et al. 2007). Feedback experts have highlighted the importance of trying to help feedback receivers to see their 'blind angles', because it is fundamental to learning to recognize the presence of a problem, to understand the nature of it and to be given specific guidance and support (Øiestad 2006). Our results indicate that medical teachers could perhaps play a more active role in identifying struggling medical students and the nature of their problems at an earlier stage, before data on final exam results are available, and before students themselves perhaps realize the full extent of their substandard performances.

Indicators of struggle

Recent research on struggling and dropout in pre-and postgraduate medical education indicates a number of signs of struggling in medical school (Table 7). It seems that *lack of participation* (e.g. absenteeism, attending late, leave of absence, social isolation, the ghost at the back present only in body, disenfranchisement from professional identity, not meeting deadlines, lack of follow-up etc.), *poor performance* (e.g. academically, clinically, professionally, failure to improve), *poor social interactions* (e.g. interpersonal problems, inappropriate interactions, poor relationships, lack of collaboration, blaming others, untrustworthy, impervious to feedback etc.), and *health problems* (e.g. depression, anxiety, psychological morbidity) could be four broad categories of signs to watch out for (Evans et al. 2010; Hayes et al. 2011; Tabby et al. 2011; Yates 2011; Mørcke et al. 2012; Roberts et al. 2012; Maher et al. 2013). Our anatomy tutors also found problems with participation, poor academic performance and poor social interactions to be signs of struggling (see

Table 6). However, in contrast to the findings in these studies our tutors also quite often mentioned signs of failing commitment amongst perceived strugglers (e.g. lack of interest, dedication, motivation, diligence etc.), whereas health problems were not identified as a perceived cause of struggling at all (Table 6). It is possible that the anatomy students sampled in this study did not feel a great need to hide their level of commitment to these particular tutors, because they knew that these tutors were students like themselves with no opportunity to affect their final grades in anatomy. It is also possible, that the lack of commitment of strugglers seemed particularly conspicuous to our student tutors, because they themselves were high performing students. These student tutors did not have access to any type of administrative information on their anatomy students' health problems, which could explain the complete absence of this type of perceived cause of struggling.

In Tinto's (1975) conceptual model of dropout in higher education, both students' academic integration (e.g. grade performance and intellectual development) and their social integration (e.g. interactions with peer groups and faculty) has the power to affect their commitments to goals and institutions, which in turn influence their dropout decisions. Our results seem to fit well with this dropout model: early tutor judgments based on observations of commitment/attendance problems, academic problems or social interaction problems *in addition to* formally documented academic problems (anatomy failure) were predictive of medical school dropout.

Limitations

This study is limited by the relatively small sample of students and tutors which corresponded to only one admission cohort, so we were unable to control for cohort effects in the models presented, which may have biased results. It could also be argued that we may not necessarily be able to extrapolate the results from one course (anatomy) in a program to other courses and learning situations. After all, anatomy is different from other courses in the medical program with its overwhelming focus on factual knowledge, where many other medical subjects (e.g. physiology and many clinical subjects) require higher order thinking or reasoning skills in addition to memorization skills. However, since anatomy courses are often scheduled quite early in many medical programs and often are a great challenge for medical students, it is probably relevant to look for early struggling in anatomy courses in many medical programs.

In addition, the generalisability of our results may be limited by the educational context in which the study took place (e.g. the type of curriculum (non-PBL), the degree of student activation in class, the number of contact hours etc.), and by the degree of tutor experience (novices versus experts). It is also a limitation that we only followed the cohort for 2.5/3 years, and not until all students had either graduated or dropped out (i.e. potentially for up to 12 years). However, attrition in this medical school tend to occur in the first year of study (Mørcke et al. 2012), and the students are required to pass the anatomy course within the first 2 years of study otherwise they are automatically expelled from the program. This means that the bias from the incomplete follow-up is less likely to be severe. Furthermore, we cannot rule out, that acute illness on the examination day could be the cause of some of the anatomy failures we observed, as absent students are simply registered as no-shows without further explanation in the data we had access to. As noted in the ethics section, we did not have permission to handle information on students' health, which is considered sensitive data. This may have resulted in some degree of misclassification or pollution of

the variable anatomy failure, which in turn could have weakened the observed association between tutors' judgements and the outcome.

Perspectives and future research

In addition to the paucity of studies on how to best identify students at risk at the earliest possible time, the literature also reveals how challenging the task of preventing failure long term is (Tekian and Hruska 2004; Devoe et al. 2007; Stegers-Jager et al. 2013), and currently there seems to be little evidence to support 'best practice' in remediation (Hauer et al. 2009; Cleland et al. 2013; Winston et al. 2014). On the optimistic side, active learning curricula have been found to improve program completion rates and times (Iputo and Kwizera 2005; Schmidt et al. 2009), and it seems that prevention and remediation strategies which encourage student participation and ownership are worth examining further (Cleland et al. 2005, Stegers-Jager et al. 2013; Winston et al. 2014; Bierer et al. 2015; Patel et al. 2015; Gonsalvez et al. 2015). The results presented in this study add evidence to the growing literature on struggling medical students, indicating that teachers should probably worry less about inducing failure by acting on their perceptions of struggling or 'opening a can of worms', as they are more often than not justly alarmed about perceived strugglers—even when their perceptions are only based on informal judgments of in-class activities in large group settings and early in a course. As the results showed, these early identified strugglers failed anatomy or dropped out as predicted by the student tutors in 3 out of 4 cases, so if anything a laissez-faire approach to struggling was arguably inimical to learning and progression. In future studies, it may be worth examining whether highly experienced academic staff in positions of power relative to the students they observe are more or less precise in identifying strugglers than for instance student tutors with less academic experience and power are.

As outlined above, only few studies on the diagnostic quality of early teacher judgments of struggling in medical education currently exist, so more studies examining this seemingly underused source of information are needed. It seems worth examining further if a combination of teachers' independent informal judgements of specific in-class behaviour (participation, commitment, academic performance, social interactions, distress) in addition to early performance testing may strengthen our ability to correctly identify students at risk of failure early. Valid early diagnosis of strugglers is the best foundation for successful early support and prevention.

References

Adam, J., Bore, M., McKendree, J., Munro, D., & Powis, D. (2012). Can personal qualities of medical students predict in-course examination success and professional behaviour? An exploratory prospective cohort study. *BMC Medical Education, 12*, 69.
Altman, D. G. (1991). *Practical statistics for medical researchers*. Boca Raton: Chapman & Hall/CRC.

Bierer, B., Dannefer, E. F., & Tetzlaff, J. E. (2015). Time to loosen the apron strings: Cohort-based Evaluation of a learner-driven remediation model at one medical school. *Journal of General Internal Medicine, 30*(9), 1339–1343.

Brenner, A. M., Mathai, S., Satyam, J., & Mohl, P. C. (2010). Can we "predict" problem residents? *Academic Medicine, 85*(7), 1147–1151.

Challis, M. (2000). AMEE medical education guide no. 19: Personal learning plans. *Medical Teacher, 22*(3), 225–236.

Cleland, J., Arnold, R., & Chesser, A. (2005). Failing finals is often a surprise for the student but not the teacher: Identifying difficulties and supporting students with academic difficulties. *Medical Teacher, 27*(6), 504–508.

Cleland, J., Leggett, H., Sandars, J., Costa, M. J., Patel, R., & Moffat, M. (2013). The remediation challenge: theoretical and methodological insights from a systematic review. *Medical Education, 47*(3), 242–251.

Devoe, P., Niles, C., Benjamin, A., Blacklock, L., Brainard, A., Colombo, E., et al. (2007). Lessons learned from a study-group pilot program for medical students perceived to be 'at risk'. *Medical Teacher, 29*(2), e37–e40.

Dunning, D., Johnson, K., Ehrligner, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*(3), 83–87.

Evans, D. E., Alstead, E. M., & Brown, J. (2010). Applying your clinical skills to students and trainees in academic difficulty. *The Clinical Teacher, 7*(4), 230–235.

Fortin, Y., Kealey, L., Slade, S., & Hanson, M. D. (2015). Investigating Canadian medical school attrition metrics to inform socially accountable admissions planning. Medical Teacher, 1–5 [Epub ahead of print].

Georg, W. (2009). Individual and institutional factors in the tendency to drop out of higher education: A multilevel analysis using data from the Konstanz Student Survey. *Studies in Higher Education, 34*(6), 647–661.

Gonsalvez, D. G., Ovens, M., & Ivanusic, J. (2015). Does attendance at anatomy practical classes correlate with assessment outcome? A retrospective study of a large cohort of undergraduate anatomy students. *BMC Medical Education, 15*, 231.

Harackiewicz, J. M., Barron, K. E., Tauer, J. M., & Elliot, A. J. (2002). Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation. *Journal of Educational Psychology, 94*(3), 562–575.

Hauer, K. E., Ciccone, A., Henzel, T. R., Katsufrakis, P., Miller, S. H., Norcross, W. A., et al. (2009). Remediation of the deficiencies of physicians across the continuum from medical school to practice: A thematic review of the literature. *Academic Medicine, 84*(12), 1822–1832.

Hayes, R. B., Lawson, M., & Gray, C. (2011). Problems presented by medical students seeking support: A possible intervention framework. *Medical Teacher, 33*(2), 161–164.

Hodges, B., Regehr, G., & Martin, D. (2001). Difficulties in recognizing one's own incompetence: Novice physicians who are unskilled and unaware of it. *Academic Medicine, 76*(10 Suppl), S87–S89.

Hojat, M., Gonnella, J. S., Erdmann, J. B., & Veloski, J. J. (1996). The fate of medical students with different levels of knowledge: Are the basic medical sciences relevant to physician competence? *Advances in Health Sciences Education, 1*(3), 179–196.

Iputo, J. E., & Kwizera, E. (2005). Problem-based learning improves the academic performance of medical students in South Africa. *Medical Education, 39*(4), 388–393.

Juul, S. (2004). *Epidemiologi og evidens*. København: Munksgaard Danmark.

Kaufman, D. M., & Hansell, M. M. (1997). Can non-expert PBL tutors predict their students' achievement? An exploratory study. *Academic Medicine, 72*(Suppl. 1), S16–S18.

Kilminster, S., Cotrell, D., Grant, J., & Jolly, B. (2007). AMEE guide no. 27: Effective educational and clinical supervision. *Medical Teacher, 29*(1), 2–19.

Kleinbaum, D. G., & Klein, M. (2002). *Logistic regression* (2nd ed.). New York: Springer.

Loyens, S. M. M., Rikers, R. M. J. P., & Schmidt, H. G. (2007). The impact of students' conceptions of constructivist assumptions on academic achievement and drop-out. *Studies in Higher Education, 32*(5), 581–602.

Maher, B. M., Hynes, H., Sweeney, C., Khashan, A. S., O'Rourke, M., Doran, K., et al. (2013). Medical school attrition-beyond the statistics. A ten year retrospective study. *BMC Medical Education, 13*, 13.

Mazor, K. M., Fischer, M. A., Haley, H. L., Hatem, D., & Quirk, M. E. (2005). Teaching and medical errors: primary care preceptors' views. *Medical Education, 39*(10), 982–990.

Menard, S. (2002). *Applied logistic regression analysis* (2nd ed.). Thousand Oaks: Sage publications Inc.

Mørcke, A. M. M., O'Neill, L., Kjeldsen, I. T., & Eika, B. (2012). Selected determinants may account for dropout risks among medical students. *Danish Medical Journal, 59*(9), A4493.

O'Neill, L. D., Wallstedt, B., Eika, B., & Hartvigsen, J. (2011). Factors associated with dropout in medical education: A literature review. *Medical Education, 45*(5), 440–454.

Øiestad, G. (2006). *Feedback*. Copenhagen: Dansk Psykologisk Forlag.

Papadakis, M. A., Hodgson, C. S., Theherani, A., & Kohatsu, N. D. (2004). Unprofessional behavior in medical school is associated with subsequent disciplinary action by a state medical board. *Academic Medicine, 79*(3), 244–249.

Papadakis, M. A., Teherani, A., Banach, M. A., Knettler, T. R., Rattner, S. L., Stern, D. T., et al. (2005). Disciplinary action by medical boards and prior behavior in medical school. *New England Journal of Medicine, 353*(25), 2673–2682.

Patel, R., Tarrant, C., Bonas, S., Yates, J., & Sandars, J. (2015). The struggling student: A thematic analysis from the self-regulated learning perspective. *Medical Education, 49*(4), 417–426.

Roberts, N. K., Williams, R. G., Klingensmith, M., Sullivan, M., Boehler, M., Hickson, G., et al. (2012). The case of the entitled resident: a composite case study of a resident performance problem syndrome with interdisciplinary commentary. *Medical Teacher, 34*(12), 1024–1032.

Schmidt, H. G., Cohen-Schotanus, J., & Arends, L. R. (2009). Impact of problem-based, active learning on graduation rates for 10 generations of Dutch medical students. *Medical Education, 43*(3), 211–218.

Stegers-Jager, K. M., Cohen-Schotanus, J., & Themmen, A. P. (2013). The effect of a short integrated study skills programme for first year medical students at risk of failure: A randomized controlled trial. *Medical Teacher, 35*(2), 120–126.

Steinert, Y. (2013). The "problem learner": whose problem is it? AMEE Guide No. 76. *Medical Teacher, 35*(4), e1035–e1045.

Stetto, J. E., Gackstetter, G. D., Cruess, D. F., & Hooper, T. I. (2004). Variables associated with attrition from uniformed services university of the health sciences medical school. *Military Medicine, 169*(2), 102–107.

Tabby, D. S., Majeed, M. H., & Schwartzman, R. J. (2011). Problem neurology residents. A national survey. *Neorology, 76*(24), 2119–2123.

Teherani, A., Hodgson, C. S., Banach, M., & Papadakis, M. A. (2005). Domains of unprofessional behavior during medical school associated with future disciplinary action by a state medical board. *Academic Medicine, 80*(10 Suppl), S17–S20.

Tekian, A., & Hruska, L. (2004). A review of medical school records to investigste the effectiveness of enrichment programs for "at risk" students. *Teaching and Learning in Medicine, 16*(1), 28–33.

Tinto, V. (1975). Dropouts from higher education: A theoretical synthesis of recent research. *Review of Educational Research, 45*(1), 89–125.

Van de Watering, G. A., & Claessens, S. (2003). Verschillen tussen de perceptie van tutoren op de leer-prestatie en de toetsprestatie van hun studenten [Differences between tutors'perceptions of perfor-mandce and students' actual achievement]. *Tijdschrift voor Hoger Onderwijs, 21*(3), 199–214.

Van den Berg, M. N., & Hofman, W. H. A. (2005). Student success in university education: A multi-measurement study of impact of student and faculty factors on study progress. *Higher Education, 50*(3), 413–446.

Whitfield, C. F., & Xie, S. X. (2002). Correlation of problem-based learning facilitators' scores with student performance on written exams. *Advances in Health Sciences Education, 7*(1), 41–51.

Wijnia, L., Loyens, S. M. M., Derous, E., Koendjie, N. S., & Schmidt, H. G. (2013). Predicting educational success and attrition in problem-based learning: Do first impressions count? *Studies in Higher Education, 38*(1), 1–16.

Winston, K. A., Van der Veluten, C. P. M., & Scherpbier, A. J. J. A. (2010). An investigation into the design and effectiveness of a mandatory cognitive skills programme for at-risk medical students. *Medical Teacher, 32*(3), 236–243.

Winston, K. A., van der Vleuten, C. P. M., & Scherpbier, A. J. J. A. (2014). Prediction and prevention of failure: An early intervention to assist at-risk medical students. *Medical Teacher, 36*(1), 25–31.

Yao, D. C., & Wright, S. M. (2001). The challenge of problem residents. *Journal of General Internal Medicine, 16*(7), 486–492.

Yates, J. (2011). Development of a 'toolkit' to identify medical students at risk of failure to thrive on the course: an explorative retrospective case study. *BMC Medical Education, 11*, 95.

Yates, J., & James, D. (2006). Predicting the "strugglers": a case-control study of students at Nottingham University Medical School. *British Medical Journal, 332*(7548), 1009–1013.

Yates, J., & James, D. (2010). Risk factors at medical school for subsequent professional misconduct: multicenter retrospective case-control study. *British Medical Journal, 340*, c2040.

Yu, T., Wilson, N. C., Singh, P. P., Lemanu, D. P., Hawken, S. J., & Hill, A. G. (2011). Medical students-as-teachers: a systematic review of peer-assisted teaching during medical school. *Advances in Medical Education and Practice, 2*, 157–172.

Zbieranowski, I., Takahashi, S. G., Verma, S., & Spadafora, S. M. (2013). Remediation of residents in difficulty: A retrospective 10-year review of the experience of a postgraduate board of examiners. *Academic Medicine, 88*(1), 111–116.