**The ‘Unskilled and unaware’ effect is linear in a real-world setting**

Marina Sawdon1, Ph.D & Gabrielle Finn1, Ph.D

1Durham University, School of Medicine and Health

**Running title:** Unskilled and unaware in a real-world setting

**Corresponding author**:

Dr. Gabrielle Finn

C143, Holliday Building

School of Medicine and Health

Durham University (Queen's Campus)

University Boulevard

Thornaby-on-Tees

TS17 6BH

Tel: 0191 334 0737

Tel: 0191 334 0301 (Anatomy dept)

Fax: 0191 334 0321

**Notes on contributors:**

**Gabrielle Finn** BSc (Hons) PhD is Lecturer in Anatomy at Durham University, School of Medicine and Health. Her research interests include anatomy pedagogy, professionalism, social identity and assessment.

**Marina Sawdon** BSc (Hons) PhD is Lecturer in Physiology at Durham University, School of Medicine and Health. Her research interests include simulation, professionalism and assessment.

**ABSTRACT**

**Context**

Self-assessment ability in medical students and practising physicians is generally poor, yet essential for academic progress and professional development.

**Objective**

The aim of this study was to determine undergraduate medical students’ ability to accurately self-assess their exam performance in a real-world, high-stakes exam setting, something not previously investigated.

**Methods**

Year 1 and Year 2 medical students (n=74) participated in a self-assessment exercise. Students predicted their exam grade (%) on the anatomy practical exam. This exercise was completed online immediately after the exam. Students’ predicted exam grades were correlated with their actual attained exam grades using a Pearson’s correlation. Demographic data were analysed using an independent t-test.

**Results**

A negative correlation was found between students’ overall predicted and attained exam grades (*P* <0.0001). There was a significant difference between the students’ predicted grades and actual grades in the bottom, 3rd and top (*P* < 0.0001), but not 2nd quartiles of participants. There was no relationship between the students’ entry status into medical school and self-assessment ability (Y1 *P* = 0.112; Year 2, *P* = 0.236) or between males and females (Y1, *P* = 0.174). However, a relationship was determined for these variables in year 2 (*P* = 0.022). The number of hours of additional self-directed learning undertaken did not influence students’ self-assessment in both years.

**Conclusion**

Our results demonstrate the ‘unskilled and unaware’ phenomenon in a real-world, high-stakes and practice-related setting. Students in all quartiles were unable to self-assess their exam performance, except for a group of mid-range students in the 2nd quartile. Poor performers were shown to overestimate their ability, and conversely, high achievers underestimated their performance. We present evidence of a strong, significant linear relationship between medical students’ ability to self-assess their performance in an anatomy practical exam, and their actual performance; in a real world setting. Despite the limited ability to self-assess reported in the literature, our results may inform approaches to revalidation which currently frequently rely on an ability to self-assess.

**INTRODUCTION**

In what is now a widely cited paper, Kruger and Dunning1 described evidence supporting the concept of ‘unskilled and unaware’. However, this was shown in experimental settings unrelated to medicine and not directly relevant to the academic ambitions of the subjects. The authors1 observed that unskilled subjects (i.e. subjects who scored low on a variety of tests) overestimated their own performance (indeed, thought themselves above average), and lacked the meta-cognitive ability to recognise that this was the case. Nor could they accurately assess the performance of others. They also found it difficult to respond effectively to feedback in this state, but where it did prove possible to improve their metacognitive ability with training, their self-assessment ratings then tended to improve. Conversely, skilled subjects tended to slightly under-rate their own performance, but this improved after feedback.

Kruger & Dunning’s study1 could have been criticised on the basis that the tasks were irrelevant and likely to have been perceived by participants as arbitrary. In particular, it could be proposed that health care learners engaged in tasks which would be relevant to their future careers might have higher degrees of engagement and more accurate perceptions of their own performance. However, a study of communication skills of family medicine residents before and after benchmarking accorded with many of Kruger and Dunning’s observations2. In addition, a study of dental students3 compared self-assessment to peer and trainer assessment in realistic environments, although the study was for research purposes only (i.e. not summative or formative in nature). They found that while peers and trainers gave similar ratings, self-assessments were over estimates, and the disparity was greatest at lowest ratings. Interestingly, they also found evidence of planned collusive behaviour, which could be summarised as “I’ll mark you high if you mark me high”.

Context independent challenges to Kruger and Dunning1 were posed by a number of authors. These included, firstly, the effect of the intrinsic difficulty of the task4. Secondly, the possibility of regression to the mean – by this logic, poor students would regress to the mean and appear to over-rate their performance, while good performers would also regress in the opposite direction, and appear to underestimate their performance5. Finally, another possibility posed was that of false reporting – that weak performers were aware of their inadequacies but were ashamed to admit them (advanced by Kruger and colleagues6 but drawing on, for example, the work of Baumeister7). Kruger and colleagues were able to provide evidence that none of these factors were significant, reinforcing their original case6, 7.

Thus, it would be valuable to provide further evidence of the ‘unskilled and unaware’ phenomenon in a relevant setting in which the assessments are summative (high stakes) and relevant to the participants’ future careers. Demonstrating this phenomenon within the medical school context is important as throughout their education, and future medical careers, medical students will be required to self-assess. This will be both on a formal and informal or self-motivated basis. Self-assessment may form one arm of a multi-source feedback exercise or could inform an individual deciding to do further reading after identifying a gap in their knowledge. Both examples require the subject to make an informed decision about their own ability to perform on some level.

Despite the consensus within the literature that ability to self-assess is generally poor8-11, self-assessment is not a wasted exercise. Articles advocate self-assessment as a key step in professional development9-11, 13. Completion of self-assessment exercises promotes reflection and thus self-awareness. This alone is beneficial to students as they continue on a path of professional and personal development. Such opinions and self-judgements will drive the students’ learning. The emotional response to completing a self-assessment, coupled with receiving feedback, is necessary for learning8. Boud8 argues that feelings and emotion are involved in learning, and that any reflection would be inhibited if “these dimensions are not accepted as integral to it.”

The aim of this study was thus to determine undergraduate medical students’ ability to accurately self-assess their exam performance in a real-world setting, something not previously investigated. The practical anatomy examination was chosen as this examination is objectively marked, high-stakes and the information taught and examined is relevant to all stages of the students’ future career.

**METHOD**

This study was approved by the School of Medicine and Health Ethics Committee.

All first and second year medical students from the Phase One (pre-clinical) Medicine Programme (199 students) at Durham University, UK were invited to participate in this research during the academic year 2009-10. Invitations were sent via email, via the Virtual Learning Environment (VLE) and posted on notice boards.

The self-assessment exercise was delivered online using the secure VLE. All students were emailed a link to the online survey. Students were able to view the exercise before deciding whether or not to participate. This exercise was available for completion immediately after the end of year (May) exam diet.

Participating students were asked to predict their grade on the Timed Practical Circuit (TPC) Anatomy exam. The Timed Practical Circuit, also known as the ‘spotter’, is a practical exam where students proceed around the laboratory answering questions at each station. Year 1 students have 30 stations, year 2 students have 60. Each station is a cadaveric specimen or anatomical model which has a feature pinned. Students must answer a question relating to the pinned structure, for example, its function.

Students were required to predict their TPC grade as a percentage. In addition, students were asked to self-report how frequently they visited the dissecting room (DR) for self-directed study (SDL) of anatomy. Students selected from 4 categories of SDL attendance; never attended, rarely attended, frequently attended, and attended the majority of SDL sessions available to them. SDL is an opportunity for students to visit the DR and consolidate knowledge, outside the usual anatomy timetabled teaching. Demographic data were also collected for each participant (entry qualification and sex).

**Statistical analysis**

The sample size was informed by previous studies14-16 conducted at our institution, showing that educational interventions achieved an effect size between 0.2-0.5 with samples sizes of 80-100 participants. Therefore, it is anticipated that the 103 potential Phase 1, Year 1 Durham invited participants will be sufficient to demonstrate an effect.

All data were analysed using SPSS® 15.0 for Windows.

For each year group, the students’ attained anatomy grades were correlated with their predicted grades and accuracy scores using Pearson’s correlation. Demographic data was analysed using an independent t-test. For both tests a value of *P* < 0.05 was considered statistically significant.

In order to establish whether the accuracy of a students’ ability to self-assess exam performance was influenced by demographics and attendance at additional sessions (self-directed learning) an accuracy score was calculated. This was determined by subtracting the predicted exam grade from the attained exam grade for each student in turn. The mean accuracy scores for males versus females were compared using an independent t-test. Any differences between accuracy scores and SDL attendance were determined using a one way ANOVA, followed where appropriate by Tukey’s *post-hoc* test. SDL attendance was grouped into 4 categories, as described earlier. Any significant difference highlighted was then subject to further analysis using an independent t-test.

**RESULTS**

Of the 97 first year students invited to participate, 31 (31%) completed the self-assessment exercise. Completion from second year students was 43 out of 102 (42%). Across both year groups, 74 out of a possible 199 students (37%) participated.

**Demographics and descriptive statistics**

There was no relationship between the students’ entry status into medical school (Graduate or High School entry) and their ability to self-assess (Year 1 *t* = 1.64, *P* = 0.11; Year 2 *t* = 1.21, *P* = 0.24). See Table 1.

There was no significant difference between males and females in year 1 and their ability to self-assess (*t* = 1.39, *P* = 0.17). However, a statistically significant difference was determined for these variables in year 2 students (*t* = -2.38, *P* = 0.02). See Table 1.

|  |  |  |  |
| --- | --- | --- | --- |
| **Demographics** | **Year 1** | **Year 2** | **Both years** |
| **Sex** | Male | 19 | 14\* | 33 |
| Female | 12 | 29\* | 41 |
| **Entry status** | Graduate entry | 14 | 15 | 29 |
| High school leaver | 17 | 25 | 42 |
| Other | 0 | 3 | 3 |

*\*P* = 0.02

**Table 1**: Demographic data for each cohort. Values indicate the number of students.

The descriptive statistics for the accuracy scores, actual and predicted grades for all participants are shown in Table 2.

|  |
| --- |
|  |
|  | N | Minimum | Maximum | Mean | Std. Deviation |
| accuracy | 74 | -66 | 68 | 6.07 | 29.421 |
| actual | 74 | 34.00 | 82.00 | 62.0811 | 10.84480 |
| pred | 74 | 10.00 | 100.00 | 56.0135 | 21.75486 |
| Valid N (listwise) | 74 |  |  |  |  |

**Table 2**: Descriptive statistics for the accuracy scores (actual grade minus predicted grade), actual and predicted (pred) grades.

**Accuracy of self-assessment**

For both year groups, a highly statistically significant, negative correlation was found between students’ predicted and attained anatomy exam grades (Year 1 *r* = -0.59, *P* <0.0001; Year 2 *r* = -0.62, *P* <0.0001, both years combined *r* = -0.58, *P* <0.0001). This is demonstrated in Figure 1. There was also a highly statistically significant positive correlation between students’ actual grades and their accuracy scores (see Figure 2; Year 1 *r* = 0.811, *P* <0.0001; Year 2 *r* = 0.808, *P* <0.0001, both years combined *r* = 0.799, *P* <0.0001).

There was a highly statistically significant correlation between students’ actual anatomy exam grades and their accuracy scores (Year 1 *r* = 0.811, *P* <0.0001, Year 2 *r* = 0.808, *P* <0.0001, both years combined *r* = 0.799, *P* <0.0001; see Figure 2).

The number of hours of additional self-directed learning that students had undertaken did not influence their ability to self-assess exam performance (Year 1 *F* = 0.69, *P* = 0.57; Year 2 *F* = 3.26, *P* = 0.05).

There was a highly statistically significant difference between the students’ predicted grades and their actual grades in the bottom (*t* = 10.58, *P* < 0.0001), 3rd (*t* = -8.31, *P* < 0.0001), and top quartiles (*t* = 14.90, *P* < 0.0001) of participants. However there was no significant difference between predicated and actual grades in the 2nd quartile (*t* = -1.82, *P* = 0.077). See Table 3 and Figure 3.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|    | **Bottom quartile****(n=18)** | **2nd quartile****(n=19)** | **3rd quartile****(n=19)** | **Top quartile****(n=18)** |
| **Mean actual grade (%)** | 52 | 57 | 67 | 73 |
| **Mean predicted grade (%)** | 83\*\*\* | 63 | 49\*\*\* | 29\*\*\* |
| **Mean accuracy score (actual minus predicted)** | -31 | -6 | 17 | 44 |

\*\*\**P* < 0.0001 mean predicted grade versus mean actual grade.

**Table 3**: The means of the actual and predicted grades, and accuracy scores (the magnitude of the difference between actual grades and predicted grades) for the quartiles of the actual grades.

**DISCUSSION**

The results presented in this study have demonstrated the ‘unskilled and unaware’ phenomenon in a real-world, high-stakes setting, relevant to students’ future career. Medical students are unable to accurately self-assess their anatomy exam performance. Poor exam performers were shown to overestimate their ability, and conversely, high achievers underestimated their exam performance (see Figures 1 & 2). Figure 2 and Table 3 demonstrate clearly that poor performers have a negative accuracy score (i.e. they overestimated their performance) and the high achieving students tended to have positive accuracy scores, meaning they underestimated their performance. However, there appears to be a cluster of students in the second quartile that are predicting their grades with a fair degree of accuracy (Table 3). This sub-set of students will be explored further in a follow up study.

This general lack of self-awareness in the students, particularly at the two extremes of ability, replicates the findings of Kruger and Dunning1. Kruger and Dunning suggest that “this overestimation occurs, in part, because people who are unskilled in these domains suffer a dual burden: not only do these people reach erroneous conclusions and make unfortunate choices, but their incompetence robs them of the metacognitive ability to realise it”1. However, Kruger and Dunning1 found the biggest discrepancy in the poor performing students; the ‘unskilled and unaware’. In this study the biggest discrepancy was seen in the high performing students (Table 3). This may, in part, be due to the reasonably high level of performance across the whole class; the mean of the bottom quartile was 52%.

The inaccuracy of self-assessment could be due to many factors; for example, inaccurate recall, modesty or arrogance. Students may have selective memories, only recalling situations where they perceive themselves to have done well. Our findings regarding academic exam performance are unique in that they are based on high-stakes assessment. However, the patterns that we have observed mirror the literature in that students are consistently poor at self-assessing9-11.

Previous research1 relied on trichotomising or treating data as quartiles. This is probably because such approaches require no inferences about the nature of the relationship between the properties under study. However, we have found evidence of a strong and significant ***linear*** relationship between medical students’ ability to self-assess their performance in an anatomy practical exam, and their actual performance, in a real world setting (Figure 1). This is also important, since Kruger & Dunning’s studies1 were originally carried out in psychology experiments in which the information being examined was not of significance to the participants. On the contrary, in this case, medicine is an extremely high stakes subject and the information under examination (anatomy) was of direct practical significance to their future careers as doctors, and their performance determines the likelihood of getting the post of their choice (or even a post at all) in the Foundation Programme after graduation.

Our data suggests that the inability to self-assess is not related to sex in year 1 students. Conversely, a moderate relationship was determined in year 2 students; females appear to be significantly more accurate at self-assessing exam performance. However, we would not wish to over emphasise this relationship due to the ratio of females to males being 2:1 within this cohort (see Table 1). Programme entry qualification status does not appear to affect the ability to self-assess either. Entry status could, in this UK context, be a proxy measure of age as all graduates are over 21 years of age and school-leavers are between 18-21 years of age.

Additionally, our finding that self-assessment ability was not linked to the number of hours of self-directed study students spent in the dissecting room appears counter-intuitive. However, these data were self-reported measures and therefore may not be accurate. Despite the potential inaccuracies related to this self-report measure, our findings appear to contradict Kruger and Dunning’s peripheral findings that as participants’ knowledge begins to increase, their metacognitive ability increases and their self-assessment ratings improve. However, we found no difference in the ability of students to self-assess in any SDL attendance category; so the ability of those who attended the majority of sessions in the DR to self-assess was no different to those who rarely attended. This warrants further investigation by using registers of attendance (not taken during this study) rather than potentially inaccurate self-reports of attendance.

Boud8 defines a self-assessment as “the involvement of students in identifying standards and/or criteria to apply to their work and making judgements about the extent to which they have met these criteria and standards.” The results presented here are indicative of students’ inability to identify the level of academic performance they have exhibited within the anatomy exam and therefore may imply that they are also unable to identify standards that are required to pass the anatomy exam.

Self-assessments are judgements based upon intrinsic and extrinsic factors. These factors might include one’s self-perception or eccentricities, a history of events, or influences from others, as well as many other unconscious factors. As educators we may be in a position to encourage students to consider what their intrinsic biases are so that we can steer students towards making assessments which are less influenced by presuppositions. In his meta-analysis of educational studies18, John Hattie, found that self-report grades had the largest influence on student performance, with an effect size of 1.44. He found that students have reasonably accurate understandings of their own levels of achievement. Minority students are more likely to be less accurate in their estimates or self-understanding of their achievement. A student’s expectations of success (which are sometime set lower than students could attain) may become a barrier for some students as they may only perform to whatever expectations they already have of their ability. From Hattie’s data is has been asserted that for this reason faculty must keep high expectations for their students in order to highlight to them what they can achieve. Hattie’s findings that students usually accurately predict their performance is encouraging. He reports that their predictions were based on their past achievement. Should these predictions be too low the subsequently often both students and faculty expectations will, on past performance, predict too low, then limits will be set on what is achievable. This is a notion worth consideration within medical education- setting the bar higher may enable students to be higher achievers in the long-run, and students should be guided to base their self-assessments upon previous achievement.

As illustrated by Eva and Regehr11, the term self-assessment has been used to describe a variety of processes which include self-directed assessment seeking, reflection, self-regulation and self-monitoring. Other studies17 described self-audit and self-rating. While Eva and Regehr11 may argue that using the term in such an all-encompassing manner is troublesome, it could also be argued that whatever the context of the assessment, the principle remains the same; by going through this process the student is prompted to make a judgement on their own abilities. This judgement may not be accurate but the path followed is one of self-monitoring, reflection and regulation. These skills are imperative for life-long learning.

Although external assessments are most frequently utilised for summative purposes and may be more reliable, it is imperative that medical students are able to self-assess. Personal and unguided reflections do not provide the learner with enough information to adequately guide their performance improvement11, and thus, it could be argued, their learning. Perhaps by coupling self-assessment with peer assessment or multi-source feedback, educators might be able to provide students with a context in which to consider their own performance. By considering the performance of others, students may begin to identify their own misperceptions about their own abilities. However, Kruger and Dunning1 demonstrated that their ‘unskilled and unaware’ participants were unable to accurately assess their peers’ performance because they lacked the metacognitive ability to do so.

Although the participating students in our study did not receive the results of the assessment exercise, making students aware of the results of a self-assessment exercise may serve to confirm that one cannot rely on self-assessment alone, as it is likely to be inaccurate. Students may therefore be encouraged to consider extrinsic sources of feedback on performance, such as faculty comments, exam grades and peer assessment when considering their performance.

An inability to self-assess in any situation could be problematic. However, this study required students to self-assess in a high-stakes and relevant environment. The students’ inability to accurately self-assess demonstrates the need for faculty to offer students guidance on self-evaluation. Such feedback and guidance could prove crucial for a successful career, as a lack of self-awareness could be problematic in the undergraduate and workplace environments.

A systematic review17 of the accuracy of physicians’ abilities to self-assess concluded that physicians have a limited ability to self-assess and that “the processes currently used to undertake professional development and evaluate competence may need to focus more on external assessment.” Despite the limited ability to self-assess reported in the literature17, our results may inform approaches to revalidation which currently frequently rely on an ability to self-assess.

Further work is needed to explore the sub-set of students who appear to have the ability to self-assess their exam performance (the mid-range of ability, 2nd quartile students). This may be achieved using qualitative methods such as focus groups as well as further quantitative studies. In order to ensure the findings of this study are transferable and generalisable the study will be extended to other topics (e.g. physiology) and to other institutions and stages of training. Furthermore, it would of great interest to establish whether self-assessment ability is context dependent i.e. whether it is more accurate in a formative or summative setting, whether it is a stable trait, and whether feedback on self-assessment improves assessor accuracy. Additionally, predictive validity is of interest, specifically ascertaining whether an inability to self-assess as an undergraduate is predictive of poor academic performance and career progression.

**REFERENCES**

[1] Kruger J, Dunning D. Unskilled and unaware of it: How difficulties in recognizing one’s own incompetence lead to inflated self-assessments. Journal of Personality and Social Psychology. 1999 77:1121-1134.

[2] Hodges B, Regehr G, Martin D. Difficulties in recognizing one's own incompetence: Novice Physicians who are unskilled and unaware of it. Academic Medicine. 2001;76:s87-89.

[3] Evans AE, Leeson MA, Petrie A. Reliability of peer and self-assessment scores compared with trainers’ scores following third molar surgery. Medical Education. 2007;41:866-872.

[4] Burson KA, Larrick RP, Klayman J. Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. Journal of Personality and Social Psychology. 2006;90:60–77.

[5] Krueger J, Mueller RA. Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. Journal of Personality and Social Psychology. 2002;82:180-188.

[6] Ehrlinger J, Johnson K, Banner M, Dunning D, Kruger J. Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. Organizational Behavior and Human Decision Processes. 2008;105:98-121.

 [7] Baumeister RF. A self-presentational view of social phenomena. Psychological Bulletin. 1982;91:3-26

[8] Boud D. Avoiding the traps: seeking good practice in the use of self-assessment and reflection in professional courses. Social Work Education. 1999;18:121-132.

[9] Eva KW, Cunnington J, Reiter H, Keane D, Norman G. How Can I Know What I Don't Know? Poor Self Assessment in a Well-Defined Domain. Advances in Health Sciences Education 2004;9:211-224.

[10] Eva KW, Regehr G. Self-Assessment in the Health Professions: A Reformulation and Research Agenda. Academic Medicine. 2005;80:S46-S54.

[11] Eva KW, Regehr G. "I'll Never Play Professional Football" and Other Fallacies of Self-Assessment. Journal of Continuing Education in the Health Professions. 2008;28:14-19.

[13] Langendyk V. Not knowing that they do not know: self-assessment accuracy of third-year medical students. Medical Education. 2006;40:173-179.

[14] McLachlan JC, Finn GM, Macnaughton J (2009) "The Conscientiousness Index: a novel tool for exploring students’ professionalism. Academic Medicine 84(5): 559-565.

[15] Finn GM, Sawdon MA, Clipsham L & McLachlan JC. (2009). Peer estimation of lack of professionalism correlates with low Conscientiousness Index scores. Medical Education 43(10): 960-967.

[16] Finn GM, White P & Abdelbagi I. (2011). The impact of color on retention of knowledge: A body-painting study within undergraduate medicine. Anatomical Sciences Education 4(6): 311-317.

[17] Davis D, Mazmanian P, Fordis M, Van Harrison R, Thorpe K, Perrier L. Accuracy of Physician Self-assessment compared with observed measures of competence. Journal of the American Medical Association. 2006;296:1094-1102.

[18] Hattie, J. (2008). Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement. Routledge, London, UK.

****

**Figure 1**: Actual anatomy exam grades versus students’ predicted grades for both years 1 & 2 combined. *r* = -0.58, *P* <0.0001

**Figure 2**: Actual anatomy exam grades (actual) versus students’ accuracy scores (accuracy) for both year 1 & 2 students combined (*r* = 0.799, *P* <0.0001). Actual grades are displayed as a percentage. Accuracy scores were calculated as actual grades minus predicted grades for each participant. Thus, data points to the right are students that performed well, but under-predicted their grade and those to the left performed poorly but predicted a high grade.

****

**Figure 3:** Predicted ability as a function of actual ability in an anatomy spotter exam.

**Conflicts of interest:**

None

**Funding:**

None

**Contributions:**

Both authors were responsible for conception of the study design, data collection, analysis and manuscript production.

.