Author’s response to reviews

Title: Assessing the preparedness of Foundation Year 1 (FY1) doctors during the transition from medical school to the Foundation Training Programme

Authors:

Athena Michaelides (athena1.michaelides@gmail.com)

Melina Mahr (melinamahr@gmail.com)

Gaurav Pydisetty (gpydisetty@gmail.com)

Jerocin Loyala (Jerocin95@gmail.com)

Version: 4 Date: 13 Jan 2020

Author’s response to reviews:

1. Author affiliations - I am unclear of the link between all the authors (who are from the University of Nicosia Medical School, Cyprus) and the South Yorkshire Foundation School. Does one or more of them also have a role in the UK?

Response: All authors have been based in South Yorkshire teaching hospitals for the last 2 years of their MD (Doctor of Medicine) degree. Added to background (line 142-144): As final year medical students currently based in teaching hospitals in the UK, and interested in working for the NHS, we aimed to assess how confident recent graduates felt in performing the tasks outlined by the GMC.

2. Response to reviewers (point 2) alludes to a 'strengths and limitations' section but no sub-titles in the manuscript exist.

Response: Strengths and limitations section included after abstract (as per journal guidelines)

3. Methods

Online survey was send to FY1 doctors in 4 hospitals in the South Yorkshire region - it is important therefore to include the total number of doctors that received it and perhaps a breakdown by site. Denominator data is one way to determine if the response rate is likely to be representative. The site breakdown would also give an indication of whether there is a bias in responses and if the 'health systems' which are asked about are unique to each site.

Response: As responded to previous reviewer: The survey was sent to the foundation school who sent out our email, so the number of responses per deanery are not traceable

4. Confidence vs competence - which is this study measuring? The conclusion and 0-6 scale suggest the former but elsewhere it suggests the latter. One could argue that medical schools are tasked with reaching a competence threshold (they work within an 'outcome based education' (OBE), or
competency based medical education (CBME) framework). This is mentioned in the background section (lines 78-80): "junior doctors are expected to be competent … from their first posting". The threshold for confidence is considered to be higher and may never be reached until students start practicing as doctors. It may seem like a philosophical point but I raise this because the terms are used interchangeably throughout the manuscript (see lines: 57, 60, 63, 141, 179, 186, 215, 232, 311, 381). This was even the case when describing the 0-6 scale items with both competence and confidence being used in the description of '0' (line 26 vs 548).

Response: There is debate across the literature between the 2 terms. We have used it as follows: Although the doctor might be competent (objectively) in that they are able to perform a task well, they might lack confidence (subjective) and perceive that they do not have this ability – for this reason we have chosen the term confidence. We are assessing the doctors’ self-belief that they can perform a certain task, so we have changed ‘competence’ to ‘confidence’ when referring to the scores.

To avoid confusion, we have used the words competence/competency/competencies as a noun to refer to skills/abilities (i.e. the competencies outlined by the GMC that they are expected to be confident in performing) [lines 169, 295, 373].

5. Survey construction - AMEE guide 87 provides an overview of how to design surveys. Whilst some steps were taken to ensure completeness and coherence there are a number of significant omissions. Firstly, given the amount of work previously published in this area (as detailed in the background section) why was there not a search of the literature to ascertain if there was a validated questionnaire that has already been used? Adaptation of an existing, peer reviewed, questionnaire would surely be the first place to start. The librarian would normally be used in this regard rather than just to check language/clarity etc. Also, what steps were taken to ensure items were interpreted correctly given that different hospitals had different 'systems'?

Response: We wanted to highlight certain competencies across 3 domains - no survey exists to assess these 3 domains using the same method (a rating scale) as we have.

For ASSISTANTSHIP & INDUCTION domain: The assistantship is a newly developed initiative hence publications are scarce. This further emphasises the importance of our results which can act as feedback to improve the assistantship in the future

For HOSPITAL SYSTEMS domain: Data published around certain competencies is old and doesn’t include doctors’ self-evaluation using online software such as the Integrated Clinical Environment and radiological platforms. We included hospital systems as a 3rd domain to assess this

For SKILLS domain: We derived skills from the the GMC's core competencies (and kept them in the categories as outlined in the document) - this is an established framework as opposed to us asking about random skills in no specific order. Furthermore, the low Cronbach value for certain categories shows discrepancy in the grouping together of skills by the GMC’s document.

As hospitals in different sites may use different systems, we aimed to keep questions general, providing examples in brackets should the responder need clarification

6. 0-6 scale - a major flaw is the scale the authors constructed. I think this also leads to confusion when interpreting the results. This was a 6 point scale, not a 5 point one: 0, 1, 2, 3, 4, 5. Therefore you might assume the midpoint would be between 2 and 3 or even 3 itself (for a perceived 1-5 scale). However, looking at the description of confidence the threshold for feeling confident is actually 4 on the scale (4 = confident, doctor has occasional doubts; 3 = adequate, doctor doubts him/herself more often than not
- which I would argue is not 'adequate' or demonstrates confidence). This is a negatively skewed scale meaning the results were interpreted incorrectly: only scores of 4 or more (together with a lower cut off of the 95% CI above 4) can be deemed to show 'confidence'. This is further complicated by varying definitions of each item on the scale (see competence vs confidence above). As such the statements in the conclusion (line 453) cannot be supported: only 11/38 of tasks have a value above 4 (with 95% CIs that do not cross below 4). I believe a median value above 4 is needed to support the statement (line 453-454): "The majority are confident in their practical skills…" (not a mean above 4).

Response: Taking into account the scores (where 4 or above shows confidence), we have changed the conclusion that previously stated ‘the majority are confident……’ to “Since confidence scores vary for different practical skills and abilities to use the hospital systems, there is room for improvement”

7. Cronbach alpha - At no point is there an explanation of why 'internal consistency' was important to calculate. The value tells you if there is 'acceptable' or 'good' internal consistency but I don't know what this means with regards to this specific survey. There is suggestion that the authors similarly struggled with this statistical test: several alphas were negative values which suggests data error, explanations of Cronbach alpha were generic rather than interpreted in the context of this specific survey (e.g. 398-399, 434-440) and an additional term ("acceptable consistency") was used in place of internal reliability (line 432). Also, when there was questionable (&lt;0.7) or unacceptable (&lt;0.5) internal consistency its significance was not mentioned when stating mean confidence scores/95% CIs.

Response: Simple explanation of why internal consistency is important has been added. Cronbach alpha values explained in relation to survey.
‘Acceptable consistency’ replaced by ‘internal reliability’

8. Inconsistent numbers - there are differences between the value in the tables and those quoted in the text e.g. lines 377, 382, 412.
Free text responses should be described in the results section and not presented for the first time in the discussion (lines 352-372). This should also be the case for the statement about the t-test statistical significance (lines 404-407).

Response: Text values amended to correspond to table. All values to 3 decimal places

9. Line 365 - misleading: "Overall, the respondents found that induction was more useful than the assistantship", this is not how it was asked in the survey. FY1s were asked, separately, if they found the induction useful and then if they found the assistantship useful. They were not asked to compare the 2 and decide which was 'more' useful. Some respondents would have deemed both to be useful.

Response: Changed to “Overall, responders found inductions and assistantships useful, although stated …“

10. Line 419-420 - "the low scores for these individual tasks skewed the final mean score for…" - actually the anomaly in this section is the mean score for IV cannulation (4.317) which was 0.73 above the next best score (3.585). So you could argue this skewed the final mean score in the opposite direction.
Response: "the low scores for these individual tasks skewed the final mean score for…” sentence has been removed.