

The validity of the Annual Review of Competence Progression: a qualitative interview study of the perceptions of junior doctors and their trainers.

Dr Rowena Viney, *Research Associate*<sup>1</sup>, Dr Antonia Rich, *Research Associate*<sup>1</sup>, Dr Sarah Needleman, *Senior Registrar in Clinical Oncology*<sup>2</sup>, Dr Ann Griffin, *Senior Lecturer in Medical Education*<sup>3</sup>, Dr Katherine Woolf, *Senior Lecturer in Medical Education*<sup>1</sup>

<sup>1</sup>Research Department of Medical Education,

UCL Medical School

Room GF/664, Royal Free Hospital

London NW3 2PF

<sup>2</sup>St Bartholomew's Hospital

West Smithfield

London EC1A 7BE

<sup>3</sup>Research Department of Medical Education,

UCL Medical School

74 Huntley Street

London WC1E 6AU

Corresponding author:

Dr Rowena Viney, [r.viney@ucl.ac.uk](mailto:r.viney@ucl.ac.uk), 020 3108 9587/07888 116 551

Competing interests:

All authors have completed the ICMJE uniform disclosure form at

[www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) and declare: all authors except Sarah Needleman

had financial support from the General Medical Council who commissioned this

research. Katherine Woolf receives a fee as educational consultant to the Membership

of the Royal College of Physicians (UK) Examination. No authors have any other

relationships or activities that could appear to have influenced the submitted work.

#### Funding:

The research was funded by the General Medical Council who were involved in designing the study, were kept informed of progress with the collection, interpretation and analysis of the data, and approved this report before submission. The researchers remained independent from the funders.

#### Ethical approval:

Ethics approval was provided by the University College London Ethics Committee, approval number 0511/011.

#### Guarantor:

Dr Rowena Viney

#### Contributorship:

Katherine Woolf and Ann Griffin designed the study in response to a tender from the General Medical Council. Rowena Viney, Antonia Rich, and Katherine Woolf analysed and interpreted the data with input from Ann Griffin and Sarah Needleman. Rowena Viney wrote the first draft. All authors revised it critically for important intellectual

content and approved the final version for publication. All authors agree to be accountable for all aspects of the work.

#### Acknowledgements:

Many thanks to the trainees and trainers who took part in this research; to Marcia Rigby, Natasha Malik, Catherine O’Keeffe and Lynne Rustecki for providing advice on recruitment and methods; and to UCL, LETB and Deanery administrative staff who helped with recruitment.

## Structured abstract

### Objectives

To investigate trainee doctors' and trainers' perceptions of the validity of the Annual Review of Competence Progression (ARCP) using Messick's conceptualisation of construct validity.

### Design

Qualitative semi-structured focus groups and interviews with trainees and trainers.

### Setting

Postgraduate medical training in London, Kent Surrey and Sussex, Yorkshire and Humber, and Wales in November/December 2015. Part of a larger study about the fairness of postgraduate medical training.

### Participants

96 trainees and 41 trainers, comprising UK and international medical graduates from Foundation, General Practice, Medicine, Obstetrics and Gynaecology, Psychiatry, Radiology, and Surgery, at all levels of training.

### Main outcome measures

Trainee and trainer perceptions of the validity of the ARCP as an assessment tool.

## Results

Participants recognised the need for assessment, but were generally dissatisfied with ARCPs, especially UK graduate trainees. Participants criticised the perceived tick-box nature of ARCPs as measuring clerical rather than clinical ability, and which they found detrimental to learning. Trainees described being able to populate their e-portfolios with just positive feedback; they also experienced difficulty getting assessments signed off by supervisors. ARCPs were perceived as poor at identifying struggling trainees and/or as discouraging excellence by focussing on minimal competency. Positive experiences of ARCPs arose when trainees could discuss their progress with interested supervisors.

## Conclusions

Trainee and trainer criticisms of ARCPs can be conceptualised as evidence that ARCPs lack validity as an assessment tool. Ongoing reforms to workplace-based assessments could address negative perceptions of the “tick-box” elements, encourage constructive input from seniors, and allow trainees to demonstrate excellence as well as minimal competency, while keeping patients safe.

## Keywords

Qualitative, medical training, medical education, assessment, validity, ARCP (Annual Review of Competence Progression), workplace based assessment

## Introduction

All doctors in training in the United Kingdom (UK) ('trainees') are assessed annually by the Annual Review of Competence Progression (ARCP). The ARCP is a formal and structured way of monitoring trainees at each stage of training. It is intended to protect patients and ensure that doctors gain suitable and sufficient experience and training to progress.<sup>1</sup> Using an electronic portfolio ('e-portfolio'), trainees collate evidence in of their learning and experience which includes evidence that they have undertaken a set number of workplace-based assessments such as direct observations of procedural skills, case note reviews, and self-reflective learning logs. The e-portfolio is reviewed against a relevant curriculum<sup>2</sup> by a panel who decide whether the trainee can progress.<sup>3</sup>

There is little research on the ARCP panel; however, workplace-based assessments have received more attention, with findings showing they are not always positively received by trainees. A narrative review of medical and dental workplace-based assessments found negativity due to assessments having unclear purpose, providing insufficient quality feedback, and a lack of time to carry them out.<sup>4</sup> Formative

workplace-based assessments are felt to be more educationally valuable than summative ones,<sup>5, 6</sup> and there are calls to move away from the tick-box culture of workplace-based assessments.<sup>7-9</sup> Opinions on specific summative workplace-based assessments are mixed: mini-clinical evaluation exercises are viewed as beneficial for development but difficult to implement,<sup>10, 11</sup> and multi-source feedback can be effective but also unhelpful.<sup>12</sup> Relatively few studies have looked specifically at ARCPs in the UK, and those few report mixed views, with some trainees finding the e-portfolio confusing and lacking in educational value,<sup>13</sup> and others feeling confident using the system and finding the panel fair but lacking individual feedback.<sup>14</sup> There are also concerns about the fairness of ARCP outcomes because doctors who graduated from medical schools outside the UK are at increased risk of poorer outcomes than UK graduates.<sup>15</sup>

The quality of an assessment or test is typically considered in terms of psychometric validity and reliability. Most assessments in medical education measure constructs<sup>16</sup> such as 'educational achievement' and 'educational ability', so all validity can be considered construct validity.<sup>17</sup> <sup>16</sup> Construct validity relates to how an assessment is constructed and administered in practice, and how its results are interpreted and used. Multiple sources of evidence are required to consider an assessment valid for use in a



specific context for a specific purpose,<sup>16</sup> of which there are five main sources (see Table 1).<sup>17, 18</sup>

**Table 1: Five sources of validity evidence based on Messick<sup>17</sup> and *The Standards for Educational and Psychological Testing*,<sup>18</sup> adapted from Cook et al<sup>19</sup> and Cook and Beckman<sup>20</sup>**

|                                      |   |
|--------------------------------------|---|
| 1) Content                           | Examines the relationship between the construct being measured and the content of the test used to do so. <sup>19</sup> |
| 2) Response process                  | Looks at the processes connecting what is being observed to the documentation of the observation. <sup>19</sup>         |
| 3) Internal structure                | The reliability of the assessment and the coherence between the assessment components. <sup>20</sup>                    |
| 4) Relationship with other variables | Considers how well the assessment correlates with other assessments testing the same construct. <sup>20</sup>           |

|                 |   |
|-----------------|---|
| 5) Consequences | Refers to the impact of the assessment on the trainee, the institution, patients, and any other affected parties. <sup>19</sup> |
|-----------------|---|

In this study we examined the validity of ARCPs by exploring how ARCPs are perceived by trainees and trainers, using Messick<sup>17</sup> and Cook et al<sup>19</sup> to guide our analysis. The data were gathered as part of a study of perceptions of the fairness of postgraduate medical training, commissioned by the GMC.<sup>21</sup>

## Method

### Participants and data gathering

During November and December 2015 96 trainees and 41 trainers were interviewed individually or in focus groups about their experiences of postgraduate medical training, by AR (health psychologist), RV (linguist), KW (academic psychologist) and SN (clinical teaching fellow and trainee). We spoke to trainees and trainers in Foundation and six specialties: General Practice, Medicine, Obstetrics and Gynaecology, Psychiatry, Radiology, and Surgery. Participants worked in London, Wales, Yorkshire and Humber, or Kent Surrey and Sussex. We asked about aspects of teaching, learning, and fairness,

with two questions on ARCPs (See Table 2). Ethical approval was provided by University College London Ethics Committee (ref: 0511/011). Information sheets explaining the research were provided to participants before agreeing to take part; written consent was obtained at focus groups and face-to-face interviews, with verbal consent obtained for telephone interviews.

**Table 2: Questions specific to ARCPs asked in interviews and focus groups**

|                 |   |
|-----------------|---|
| <b>Trainees</b> | <p>We are particularly interested in assessments, including ARCPs and Royal College examinations.</p> <ul style="list-style-type: none"> <li>• What comes into your head when I say “ARCP”?</li> <li>• How fair do you think ARCPs are?</li> <li>• Anyone failed an exam or an ARCP? (prompt: if exam, was it written or clinical)</li> <li>• Why do you think you failed?</li> <li>• Do you think failing affected you in any way? (if necessary: How?)</li> </ul> |
| <b>Trainers</b> | <p>We are very interested in assessments, including ARCPs and Royal College examinations.</p>   |

|  |   |
|--|---|
|  | <ul style="list-style-type: none"><li>• What comes into your head when I say “ARCP”?</li><li>• How fair do you think ARCPs are?</li></ul> |
|--|---|

## Analysis

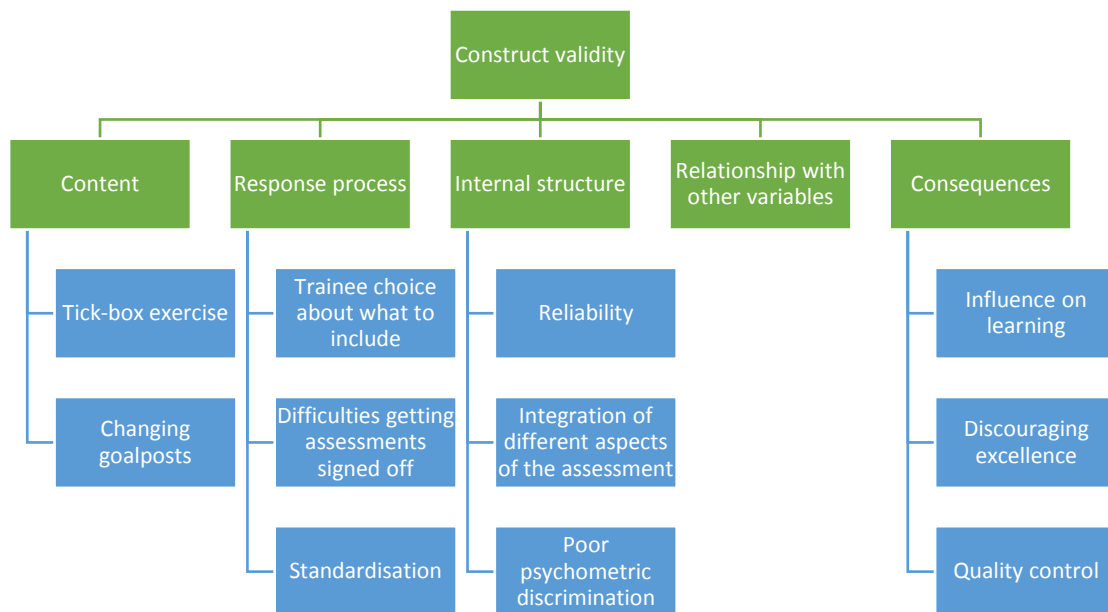
Data were transcribed professionally. The research team (RV, AR and KW) examined the data to identify emerging themes, using thematic analysis<sup>22</sup> guided by Mountford-Zimdars et al’s<sup>23</sup> analytic framework. A final coding framework was refined by KW, AR and RV after discussion. The whole data set was coded by RV, with portions of the data-set second-coded by the rest of the research team; consistency was ensured by discussing the framework with all team members and agreeing descriptors for each code, and coding discrepancies were resolved through discussion. Coding was conducted using QSR International’s NVivo 10© software.<sup>24</sup> Two primary themes emerged around ARCPs: “ARCPs are fair” and “ARCPs are not fair”. On further examination further subthemes emerged around why ARCPs were fair or unfair; these concerned the validity of ARCPs, and were then analysed using Messick’s<sup>17</sup> framework as a guide (see figure 1).

## Results

Trainee and trainer perceptions of the validity of ARCPs are presented according to the five main sources of validity evidence described in Table 1. Subthemes are shown in Figure 1. Although the data for this analysis are participants' perceptions and experiences of ARCPs, the various themes raised relate all five sources of validity evidence and not just participants' engagement with the process and its consequences.

Overall there was general dissatisfaction with ARCPs, especially among UK graduates – international medical graduates (IMGs) were more positive. Trainers tended to view the process more positively although they did voice negative views.

Figure 1: Subthemes within the five sources of construct validity evidence



## Content

### *'Tick-box exercise'*

ARCPs were described as a 'tick-box exercise' in 27 of the 65 interviews and focus groups<sup>1</sup>; this was generally a criticism of populating the e-portfolio. ARCPs were felt to test clerical ability rather than clinical ability, which some believed were inversely correlated:

---

<sup>1</sup> Based on a text search in the NVivo database for: tickbox OR tick-box OR "tick-box" OR "box ticking" OR box-ticking.

I'd say that people I've found who are very good at filling all their logs and do extremely well in the e-portfolio are actually the ones who are not very good clinically.

**Trainee/GP/ST1-3/UKG/white/female<sup>2</sup>**

Many trainees felt that competence should not be a function of the number of times a trainee has performed a procedure but whether they can perform it unsupervised.

Another common criticism was that the competencies trainees were assessed on were irrelevant to their current or future work and a waste of time:

With the practical, procedural skills that we have to get done, I feel like they're more a tick-box exercise, and they're not actually that useful because a lot of the skills are becoming more done by Radiology.

**Trainee/Medicine/ST1-3/UKG/white/female**

---

<sup>2</sup> Quotes are attributed to participants using the following coding key: Trainee/Trainer, Specialty, Grade (if trainee), UK or international medical graduate, self-reported ethnicity, self-reported gender e.g. Trainee/GP/ST1-3/UKG/white/female.

Abbreviations used are: BME (Black & Minority Ethnic), GP (General Practice), IMG (international medical school graduate), O&G (Obstetrics & Gynaecology), ST1-3 (specialist training grades 1 to 3), ST4+ (specialist training grades four or higher), UKG (UK medical school graduate).

Conversely, a few trainees and trainers felt that ARCPs covered a wide range of skills and competencies and thus provided a good sense of overall ability:

I think it is fair because it looks at ‘Have they passed the exams? ... Have they done their workplace based assessment?’ ... So I think it is fair, it does look at a large aspect of a broad training scheme.

**Trainer/GP/UKG/white/male**

#### *Changing goalposts*

Some trainees described assessment criteria changing with little or no notice. One trainee said that after completing half of the necessary workplace-based assessments she discovered that the criteria had changed and the assessments she had completed were now redundant **[Trainee/Medicine/ST4+/UK/white/female]**. In another instance, a trainee reported that miscommunication caused an entire cohort to fail for having an incorrect number of supervisor reports

**[Trainee/Medicine/ST4+/UKG/white/female]**.



Response process

*Trainee choice about what to include*

There was a perception that trainees could exclude anything negative from their e-portfolio. For example, trainees can carefully choose seniors to sign them off or give positive feedback because the trainee fits in socially:

All my [Case Based Discussions], everything has been from registrars who have generally said, 'Yeah, I'll just do one for you'. It's not been a formalised thing. It's basically been the same as the rugby tie, but rather than wearing a tie, I've just known them and get on with them, and then they'll do the thing for me.

**Trainee/GP/ST1-3/UKG/white/male**

*Difficulties getting assessments signed off*

Several trainees described difficulties getting seniors to sign off assessments due to lack of engagement or system difficulties. Other trainees were unable to complete all the necessary supervised procedures, either due to the unavailability of clinical opportunities, or because they were deemed competent to carry them out unsupervised. In extreme cases the failure of supervisors to sign off affected ARCP outcomes:

Trainee 1 I've sat in an ACRP and been told I've not got enough assessments signed, I've just pulled out the list of all the tickets that I'd sent out that hadn't been completed by consultants. ... You can't defend yourself in that situation. ...

Trainee 2 Yeah, the assumption is that the lack of work-based assessments is a reflection on the laziness or the lack of motivation or the lack of-

Trainee 1 'Failure to engage with the portfolio' is the phrase they use here.

**Trainee 1/Medicine/ST4+/UKG/white/male**

**Trainee 2/Medicine/ST1-3/UKG/white/female**

### *Standardisation*

IMGs were most likely to say that 'ticking boxes' resulted in a standardised approach making the process fairer, although other trainees felt that standardisation did not make up for a lack of content validity:

I think [ARCPs] are fair in the sense that they are a piece of standardised paperwork which anyone can learn to get filled out. I think if you ask 'Are they an effective assessment of any practical measure of doctoring ability?', that one I'm probably less certain about their quality.

## **Trainee/Medicine/ST4+/UKG/BME,/male**

### Internal structure

#### *Reliability*

There was concern at lack of consistency in ARCPs across specialties, regions, and training grades. One trainee described completing extra documentation implemented by her training programme director, which was not used in other sub-specialties or regions **[Trainee/Medicine/ST4+/UKG/white/female]**. Some voiced concern about panel reliability and fairness. For example, one trainee described how a black colleague received a lower outcome and more “hassle” than another equally experienced white trainee, which he thought was due to their ethnicity

**[Trainee/Surgery/ST4+/UKG/BME/male]**.

Several GP trainers voiced frustration at the apparent disconnect between their assessment of a trainee and that of the panel. GP trainees attend panel only if their progress raises concerns, which was compared to “Sending them to the headmaster’s office for a telling off” **[Trainer/GP/UKG/white/male]**. Yet if the panel passed a trainee, this undermined the trainers and could damage the trainer-trainee relationship irrevocably **[Trainer/GP/IMG/BME/female]**.

### *Integration of different aspects of the assessment*

Some trainees felt that the ARCP's different criteria were poorly integrated: trainees could fail an ARCP for something trivial while significant achievements were ignored (see also below *Consequences: Discouraging excellence*):

... it had been a big achievement and it sort of felt like in any process that's supposed to be about your achievements and what you've done it's just completely bonkers that they hadn't mentioned that and they had mentioned the really basic thing that I hadn't got a tick-box signature to do.

**Trainee/O&G/ST4+/UKG/white/female**

### *Poor psychometric discrimination*

Several trainees felt that ARCPs were poor at discriminating between trainees of different abilities:

They're not fit for purpose because they don't identify poorly performing trainees.

They don't identify excellent trainees.

**Trainee/Medicine/ST4+/UKG/white/male**

## Relationships with other variables

Only one participant spoke about the relationship between ARCPs and other assessments. An IMG trainee knew two IMG colleagues who progressed well in their GP training (presumably passing their ARCPs) but failed their GP exit exam

**[Trainee/Psychiatry/ST4+/IMG/black/male].**

## Consequences

### *Influence on learning*

In general trainees did not feel that populating their e-portfolios encouraged learning; instead completing a large number of assessments impeded learning, either by demotivating trainees as “the more you do the less value you attach to each one”

**[Trainee/Surgery/ST4+/UKG/BME/male]**, or by taking up time that could be better spent on another educational activity:

Our e-portfolio seems to be ever expanding and sprawling and it gets to the point you wonder what the actual benefit of it is from an educational point of view. You find that you spend more time filling in boxes than you do reading about a subject.

**Trainee/O&G/ST4+/UKG/white/female**

.

There were mixed views about attending ARCP panels. Several hospital medicine trainees disliked attending panels, finding them stressful whereas others would prefer to attend the panel and have the opportunity to raise issues and get individualised feedback.

A few trainees commented that trainers' understanding of the system and their willingness to engage with it influenced how useful the ARCP process was for learning:

It all boils down to who's your supervisor and whether they understand the system, whether they're committed to you as a trainee.

**Trainee/Medicine/ST1-3/UKG/white/female**

IMGs were more likely than UK graduates to speak about the ARCP as a supportive mechanism to ensure trainees are ready for their post-training roles:

Obviously people need to make sure that you are where you should be. That's just it, they're just trying to ensure that you're getting the support you should be getting.

**Trainee/Psychiatry/ST4+/IMG/BME/female**

Trainers were generally more positive than trainees. Some had been active in improving ARCPs to be more supportive and useful for trainees, in addition to its role of checking trainees' progress.

### *Discouraging excellence*

Some trainees felt that ARCPs encouraged minimal competency at the expense of excellence, and that trainees could effectively be penalised for being competent when they started a placement:

There's the expectation to show development through the year, so you're supposed to start off bad and end up better. But if you start off good you're in real trouble.

**Trainee/Medicine/ST4+/UKG/white/male**

### *Quality control*

Quality control was seen as an important purpose of ARCPs by trainees and trainers, but the concerns described above made trainees and some trainers question whether ARCPs were able to prevent poor trainees progressing and protect patients.

## Discussion

### Statement of principle findings

Many trainees and trainers felt that ARCPs could be useful, and that assessment is necessary to check progress; however, the way that ARCPs are currently conducted is problematic. Viewed in terms of psychometric validity, participants' – especially trainees' – views suggested a lack of evidence for the validity of ARCPs as a means of assessing progress. In particular there was poor evidence for the e-portfolio's content validity with its 'tick-box' nature viewed as assessing trainees on clerical rather than clinical ability and concerns that trainees could select only positive assessments for their e-portfolio. Other major concerns were that ARCPs encourage minimal competency instead of excellence while not being sensitive enough to identify poorly performing trainees, and that ARCPs discourage learning, and disengage trainees. Attending the panel could be stressful but also an opportunity to gain individualised feedback. Positive experiences of ARCPs arose when trainees could discuss their progress with interested supervisors. IMGs felt more positive about having standardised boxes to tick which they felt was fair.



### Strengths and weaknesses of the study

This was a large study across four regions in England and Wales, involving doctors from six specialties plus foundation, from all stages at trainee or trainer level. Participants included female and male UK and international medical graduates from various ethnic backgrounds. The scale of the research resulted in a rich qualitative data-set with 137 participants from across the UK. However, we spoke to more GPs and fewer radiologists, limiting our ability to examine differences between specialities. The research was conducted during negotiations between the British Medical Association and the UK government regarding the new junior doctors' contract in England; however, the negative opinions expressed by participants reflect those reported in earlier research<sup>25, 26</sup> and so the political climate did not appear to influence participants' reports unduly. Response bias is possible, as participants volunteered in response to circulated information about the research and those with negative experiences may have been more interested in taking part; however, most participants shared both negative and positive experiences.

### Strengths and weaknesses in relation to other studies

The findings presented here reflect prior research on trainees' opinions about ARCPs and workplace-based assessments; for example, that trainees feel disillusioned with

the process and ARCPs discourage excellence,<sup>26</sup> and that ARCP outcomes are not a useful evaluation measure of a curriculum.<sup>25</sup> Our research suggests that “tick-boxes” are often perceived as reductionist and that assessments which provide more quality formative feedback during training as well as at the annual review would be beneficial. Much previous research has focused on limited geographical areas or specific specialties;<sup>13, 14, 27</sup> as this study involved participants from across England and Wales, and across specialities, its findings may have a greater reach.

Meaning of the study: possible mechanisms and implications for clinicians or policymakers

In September 2016 Health Education England acknowledged that the “tick box culture of the ARCP” has become problematic and announced a review of ARCPs to begin in October 2016.<sup>28</sup> Similarly, the Joint Royal College of Physicians Training Board has outlined the move within UK internal medicine training to a model of entrustable professional activities,<sup>29</sup> with the outcome of training being that trainees are “trusted to undertake all the key critical tasks needed to work as a consultant”.<sup>7</sup> To make a trust judgement, supervisors will need a holistic view of trainees’ abilities, and our findings suggest that the input from trainers required to accurately form this holistic view (such as frequent formative assessment) is something trainees would value. Indeed, in

related work we found that good relationships with trainers are a key influence on trainees' learning,<sup>21</sup> a point echoed in a call to return to an apprenticeship model in surgery training.<sup>30</sup> More opportunity for constructive feedback from the ARCP panel could be similarly beneficial. The 2008 Tooke report<sup>31</sup> emphasised the need for excellence in selection into postgraduate training, and our findings suggest that the revisions to workplace-based assessments should similarly allow trainees to demonstrate excellence as well as minimal competency, while keeping patients safe.

#### Unanswered questions and future research

Some participants mentioned differences in the ARCP process across specialties, grades, and regions; however, as ARCPs were not the sole focus of the research from which this paper stemmed, we were unable to include more detailed questions on this. It would be useful to establish if there are differences along these lines, and if so to investigate what they are and why they exist. Further work on the ARCPs' components, including the panel, would also be of interest, as from our data it appears that different types of workplace-based assessment are prone to different problems. With the anticipated changes to workplace-based assessments<sup>7</sup> and a call to improve the training system for trainees, including strengthening the trainee-trainer relationship,<sup>30</sup>

it will be useful to monitor trainees' and trainers' perceptions of the ARCP (or its replacement) after this change.

### Acknowledgements

Many thanks to the trainees and trainers who took part in this research; to Marcia Rigby, Natasha Malik, Catherine O'Keeffe and Lynne Rustecki for providing advice on recruitment and methods; and to UCL, LETB and Deanery administrative staff who helped with recruitment.

### Funding statement

This research into the fairness of medical postgraduate training in the UK was funded by the General Medical Council, and the report on the full findings<sup>21</sup> can be found here <http://www.gmc-uk.org/about/research/23658.asp>. The current paper describes the research participants' perceptions of ARCPs, which is one part of the findings that emerged from this project. We did not explicitly plan at the project inception to analyse data about the perceived validity of ARCPs other than in relation to differential attainment; however during the course of the study it became apparent that trainee

dissatisfaction with ARCPs was a significant finding, hence our deciding to conduct additional analyses on that section of the data.

#### Declaration of competing interests:

All authors except Sarah Needleman had financial support from the General Medical Council who commissioned this research. Katherine Woolf receives a fee as educational consultant to the Membership of the Royal College of Physicians (UK) Examination. No authors have any other relationships or activities that could appear to have influenced the submitted work.

#### References

1. The UK Foundation Programme Office. Guide to the Foundation Annual Review of Competence Progression (ARCP) Progress - 2016. 2016.
2. Black D. Revalidation for trainees and the annual review of competency progression (ARCP). *Clinical Medicine*. 2013; 13: 570-2.
3. National Health Service. The Gold Guide: A Reference Guide for Postgraduate Specialty Training in the UK. 6th ed. 2016.
4. Massie J and Ali JM. Workplace-based assessment: a review of user perceptions and strategies to address the identified shortcomings. *Advances in Health Sciences Education*. 2016; 21: 455-73.
5. O'Leary D, Al-Taiar H, Brown N, Bajorek T, Ghazirad M and Shaddel F. Workplace assessment in crisis? The way forward. *BJPsych Bull*. 2015.
6. Rees CE, Cleland JA, Dennis A, Kelly N, Mattick K and Monrouxe LV. Supervised learning events in the Foundation Programme: a UK-wide narrative interview study. *BMJ Open*. 2014; 4.
7. Black D. An end to box ticking: an overhaul of competency based education. *BMJ Careers*. 2016.

8. Talbot M. Monkey see, monkey do: a critique of the competency model in graduate medical education. *Medical Education*. 2004; 38: 587-92.
9. Miller A and Archer J. Impact of workplace based assessment on doctors' education and performance: a systematic review. *BMJ*. 2010; 341: c5064.
10. Castanelli DJ, Jowsey T, Chen Y and Weller JM. Perceptions of purpose, value, and process of the mini-Clinical Evaluation Exercise in anesthesia training. *Canadian Journal of Anesthesia-Journal Canadien D Anesthesie*. 2016; 63: 1345-56.
11. Brazil V, Ratcliffe L, Zhang JZ and Davin L. Mini-CEX as a workplace-based assessment tool for interns in an emergency department - Does cost outweigh value? *Medical Teacher*. 2012; 34: 1017-23.
12. Brown JM, Lowe K, Fillingham J, Murphy PN, Bamforth M and Shaw NJ. An investigation into the use of multi-source feedback (MSF) as a work-based assessment tool. *Medical Teacher*. 2014; 36: 997-1004.
13. Vance G, Williamson A, Frearson R, et al. Evaluation of an established learning portfolio. *The Clinical Teacher*. 2013; 10: 21-6.
14. Eynon-Lewis A and Price M. Reviewing the ARCP process: experiences of users in one English deanery. *BMJ Careers*. 2012.
15. Tiffin PA, Illing J, Kasim AS and McLachlan JC. Annual Review of Competence Progression (ARCP) performance of doctors who passed Professional and Linguistic Assessments Board (PLAB) tests compared with UK medical graduates: national data linkage study. *BMJ*. 2014; 348: g2622.
16. Downing SM. Validity: on the meaningful interpretation of assessment data. *Medical Education*. 2003; 37: 830-7.
17. Messick S. Validity. In: Linn RL, (ed.). *Educational Measurement*. 3rd ed. New York, NY: American Council on Education and Macmillan, 1989, p. 13-103.
18. American Educational Research Association, American Psychological Association and National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association, 2014.
19. Cook DA, Kuper A, Hatala R and Ginsburg S. When Assessment Data Are Words: Validity Evidence for Qualitative Educational Assessments. *Academic medicine : journal of the Association of American Medical Colleges*. 2016; 91: 1359-69.
20. Cook DA and Beckman TJ. Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application. *The American Journal of Medicine*. 2006; 119: 166.e7-e16.
21. Woolf K, Rich A, Viney R, Rigby M, Needleman S and Griffin A. Fair Training Pathways for All: Understanding Experiences of Progression. London: General Medical Council, 2016.
22. Braun V and Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology*. 2006; 3: 77-101.
23. Mountford-Zimdars AS, D; Moore, J; Sanders, J; Jones, S; Higham, L. Causes of Differences in Student Outcomes,. Higher Education Funding Council for England, 2015.

24. QSR. NVivo qualitative data analysis Software Version 10. Version 10 ed. <http://www.qsrinternational.com/>; QSR International Pty Ltd., 2012.
25. Laskaratos F-M, Gkotsi D and Panteliou E. A critical review of the core medical training curriculum in the UK: A medical education perspective. *JRSM open*. 2014; 5.
26. Dormandy L and Laycock K. Triumph of process over practice: changes to assessment of physicians. *BMJ Careers*. 2015.
27. Goodyear H, Wall D and Bindal T. Annual review of competence: trainees' perspective. *The Clinical Teacher*. 2013; 10: 394-8.
28. Reid W. Transforming postgraduate medical training in the NHS. *BMJ Careers*. 2016.
29. ten Cate O. Nuts and Bolts of Entrustable Professional Activities. *Journal of Graduate Medical Education*. 2013; 5: 157-8.
30. Lavelle-Jones M. Letter to Secretary of State for Health Jeremy Hunt. <https://www.rcsed.ac.uk/media/414502/hunt.pdf>; Royal College of Surgeons of Edinburgh, 2016.
31. Tooke J. Aspiring to Excellence: Final Report of the Independent Inquiry into Modernising Medical Careers. London: MMC Inquiry, 2008.