

Article type: Review

Assessing professional competence: a critical review of the Annual Review of Competence Progression

Katherine Woolf, Michael Page, Rowena Viney

Research Department of Medical Education

UCL Medical School

Accepted for publication in the Journal of the Royal Society of Medicine.

Summary

The Annual Review of Competence Progression (ARCP) is the process used to determine whether trainee doctors in the United Kingdom are safe and competent to progress to the next training stage. Criticisms from doctors led to a review of the ARCP in 2017, for which the current review was commissioned. The aim was to provide an evidence base to inform recommendations that would enhance the validity of the summative and formative elements of the ARCP.

Evidence obtained from a systematic search of the peer reviewed and grey literature was synthesised with information from national ARCP guidance documents, and local and specialty-specific ARCP guidance. Evidence from the wider literature on assessing competence in medical education was used to critically evaluate the summative and formative elements of the ARCP.

Trainees and trainers were skeptical about the ability of the ARCP to differentiate between different levels of professional competence, and were concerned that the ARCP only reliably identifies the most poorly performing trainees, and not trainees performing at or just under the borderline. National ARCP guidance lacked detail, resulting in variability across locations and specialties in how the ARCP is enacted, threatening the validity and reliability of outcomes. Feedback is not routinely provided to all trainees, which can leave those with performance difficulties unsupported, and high performers demotivated. Variability in the provision and quality of feedback after ARCP panels, and in preparation for panels, can negatively affect learning.

The ARCP functions as a high-stakes assessment, likely to have a significant impact on patient care. To ensure it is fit for purpose the ARCP should be subject to the same rigorous evaluation as other high-stakes assessments; there should be consistency in ARCP procedures across locations, specialties and grades; and all trainees should receive high-quality feedback.

Keywords:

assessment, competence, summative, formative, ARCP

Introduction

The Annual Review of Competence Progression (ARCP) is a yearly review of United Kingdom (UK) trainee doctors' performance against curricular milestones.¹ The ARCP is a competency-based review of whether a trainee doctor is suitable to progress to the next stage of, or to complete, their training. It aims to protect patients and the public by determining whether a trainee is safe to practice, thereby providing the mechanism by which trainees revalidate with the General Medical Council (GMC) and maintain their license to practise.¹ Thus, it is a high-stakes assessment for trainees, as an unsatisfactory ARCP can result in trainees undertaking more training, having their training time extended, or being released from their training programme.¹

Definitions and terms of reference

In this paper, we regard the ARCP process as including the systems and practices of collecting and presenting evidence about trainees' progress during their training and the judgments made about that evidence by a panel of assessors (summative elements); and any feedback given to trainees before and after the panel (formative elements). Aspects of the ARCP, notably e-portfolios and workplace-based assessments (WBAs), have been much researched (e.g.²⁻⁷), as has the competency-based model of medical training from which the ARCP arises (e.g.⁸⁻¹²). We did not set out to conduct a review of each of these constituent elements; rather, the main focus of this review is the ARCP itself. In doing so, our aim was to draw together evidence about the process and outputs in a manner that has not yet been done despite the important role that the ARCP plays in trainee progression.

Rationale for the research

Dissatisfaction with the ARCP has been recorded in the literature (e.g. Viney et al¹³), reflecting widespread unhappiness with the process that led to a review being undertaken by Health Education England (HEE) in 2017.¹⁴ This article is drawn from the research underpinning that review, which was commissioned by HEE in order to support evidence-based recommendations for improving the validity of the summative and formative elements of the ARCP.¹⁵ Thus, we present a critical review of the validity of the summative and formative elements of the ARCP obtained by synthesising information from the ARCP literature, from policy and guidance documents, and from the wider literature on assessing competence in medical education.

Methods

We performed a systematic search of Medline and PubMed databases using the search terms “ARCP”, “Annual Review of Competence Progression” and “Annual Review of Competency Progression” from January 2005 to August 2017. Backwards and forwards citation searches of relevant articles were conducted. As the ARCP’s implementation is guided by policy documents unlikely to be retrieved through database searches, we also conducted targeted searches of key policy-makers’ websites, including HEE, the GMC and the UK Foundation Programme (UKFP), for relevant reports and policy documents. The inclusion criteria were: articles or reports containing qualitative or quantitative information about ARCP process or outcomes; articles published in English. Articles or reports mentioning the ARCP but not containing information about the ARCP process or outcomes were excluded.

In analysing the literature, we took validity to mean the purpose of the ARCP, whether in its current form it is fit-for-purpose, and the extent to which it achieves its purpose.¹⁶ We reviewed and synthesised the information obtained from the search with information obtained

from official ARCP policy and guidance documents, and a sample of policy and guidance documents which adapt national policy for use in a particular location, within a particular specialty, or at a particular training grade. We critically reviewed the information obtained about the ARCP in light of evidence from the wider literature on assessing competence in medical education.

Results

Systematic search results

Searches of databases and journals plus backward- and forward-citation searching gave 297 hits in addition to 11 potentially relevant reports of which we were already aware as researchers in this field. After de-duplication, irrelevant reports were removed by screening titles and abstracts, and then by reading full-texts. This resulted in 30 reports for inclusion in the review, 17 of which were peer-reviewed. Included reports mostly related to summative aspects of the ARCP. See Supplementary Table 2 and Supplementary Figure 1 for details.

Policy information and guidance

We extracted policy information about the stated purposes of the ARCP and practical guidance on undertaking ARCPs from the official national guidance.^{1, 17-20} We also reviewed a convenience sample of local and specialty-specific ARCP guidance obtained via online searches (see Supplementary Table 1).

Summative aspects of the ARCP

Criteria and evidence used by panels to make decisions

The national ARCP guidance lacks detailed information about which evidence should be assessed and how different types of evidence should be weighted. Qualitative research

suggests this lack of clarity and the subsequent variability of evidence required by panels can undermine panels' ability to make valid, reliable decisions; it can also affect trainers' ability to guide trainees effectively through a changing curriculum while remembering the different requirements for trainees at different levels.^{13, 21, 22} To address issues of this sort, additional guidance has been issued by Medical Royal Colleges, Local Education and Training Boards (LETBs), Deaneries, and Trusts (see Supplementary Table 1); however, trainees believe that the resulting inconsistencies in requirements between different specialties, grades and regions is unfair.¹³

The Gold Guide¹ implies that the Educational Supervisor (ES) Report should take precedence over other evidence. The ARCP literature suggests that, in practice, the ES Report and the number of WBAs a trainee has recorded are both important, although their relative weighting may vary^{13, 23-26}. Different aspects of the same type of evidence may also be weighted differently - for example, two studies in General Practice suggested panels value the quality of individual WBAs over the quantity recorded,^{27, 28} whereas a small study of paediatric trainees found panels were satisfied when trainees achieved sufficient numbers of assessments and did not penalise trainees whose WBAs contained poor quality reflections.²⁹

There is evidence from the medical education literature that to make valid judgements, panels should consider a range of evidence reflecting different aspects of performance rated by a variety of assessors,³⁰ and both the number of assessments and the performance of the trainee are important.⁹ The literature is unclear about how any weighting should be applied, however greater consistency between ARCP panels would provide equitability for trainees in different specialties.

Attendance of the supervisor and the trainee at the ARCP panel

The Gold Guide¹ and the Guide to the Foundation ARCP¹⁷ take slightly different positions on the presence of ES at panel meetings. The Gold Guide states the ES should remove themselves if it is anticipated that their trainee will get an unsatisfactory outcome (implying they can be present otherwise) whereas the Foundation Programme guidance states the ES should not take part in their trainee's panel at all.

Attendance at the panel can present challenges for the ES: a supervisor in Rothwell's study described it could cause problems for their educational relationship with the trainee, particularly if the ARCP outcome is negative²¹; however another study found trainers perceived this as more problematic than trainees did.³¹ This issue is much discussed by van der Vleuten and colleagues.^{11, 30, 32, 33} While they believe supervisor input is important to increase the accuracy of panel judgements,³³ they are concerned that the crucial trainee-supervisor relationship may be compromised when the supervisor makes high-stakes decisions about trainee progression. In this regard, the Foundation Programme Guidance is better aligned than the Gold Guide with the views of leading medical education researchers.

The Gold Guide and Foundation guidance state trainees may be present at ARCP panels, and in some circumstances may even be expected to attend, such as when receiving notification of an unsatisfactory outcome. The guidance is clear however that a trainee's attendance should not contribute to panel decision-making; however, the literature confirms that when trainees do attend panels, they can feel that their attendance affects the outcome they receive.^{13, 21, 25}

Identifying poor performance and patient safety issues

The validity of the ARCP as a summative assessment hinges on its ability to reliably distinguish between satisfactory and unsatisfactory performance, between different levels of unsatisfactory performance, and to identify patient safety issues.

The ARCP literature shows trainees and supervisors have concerns that the ARCP measures clerical rather than clinical ability; that it does not reliably identify anything other than extremely poor performance; and that it cannot reliably identify patient safety concerns.^{13, 21} We found no quantitative studies linking ARCP outcomes with patient safety or fitness to practise outcomes. Six studies found ARCP outcomes are correlated with performance in other assessments,^{27, 34-38} which suggests it can distinguish between different levels of performance, although there is not currently sufficient data to know how sensitive it is. A recent review of WBAs reached a similar conclusion;² however a number of studies have found a link between examination performance and sanctions^{34, 35, 39} so it should be possible to establish whether ARCP outcomes also predict patient safety risks or other professional difficulties, particularly if the quality of ARCP data collection were to be improved.³⁸

The Gold Guide states panels will require additional information about trainees anticipated to receive a poor outcome. This may enhance the accuracy of borderline judgements; however the system relies on concerns being easy to detect, raise, and investigate, and the ‘failure to fail’ phenomenon suggests trainers may be unwilling to identify trainees who are struggling.^{21, 40} The new Generic Professional Capabilities Implementation Guidance⁴¹, jointly produced by the Academy of Medical Royal Colleges and the General Medical Council (GMC), goes some way to recognise and address the ‘failure to fail’ problem, stating supervisors should be given ‘time, training and support and be empowered to act if trainees are judged not to be making satisfactory progress’(p.13). A

recent review on ‘failure to fail’⁴⁰ supports training to overcome the problem, and also emphasises the need for ‘strong assessment systems with established criteria’ (p.1097) and ‘opportunities for trainees after failing’(p.1098). Reducing the weighting of the ES report in panel decision-making may also help,⁴⁰ and it may be helpful to formally review trainees’ progress well before the ARCP, when the stakes are lower. This might take the form of an interim or pre-ARCP panel, which is considered in the formative section below.

Reliability of the ARCP

We found no published numeric estimates of the reliability of ARCP outcomes or about the number and success rates of appeals. Our own descriptive analysis of ARCP outcome data published by the GMC showed the proportion of unsatisfactory ARCP outcomes varies by specialty and region (see Figure 1), reflecting qualitative reports from trainees that the requirements of ARCP panels vary ‘across specialties, regions and training grades’ (p.113).¹³ Without a large-scale, longitudinal multilevel analysis, it is not clear how much these differences are due to variability in factors such as trainee ability, curricular requirements, or panel decision-making approaches.

Figure 1 about here

Much of the ARCP is based around WBAs, which inform the ES Report and are presented to the panel in the e-portfolio. WBAs typically have low reliability³ but reliable judgements about a trainee’s overall competence can be made using WBAs^{6, 11} so long as there are a large number of assessments and narrative reports sampled across curriculum areas and assessors.^{30, 42} The literature suggests however that WBAs are not always well sampled because trainees have difficulty collecting evidence, or because the high-stakes nature of the ARCP means trainees are incentivised to select assessors or cases that show them in a positive light (e.g.¹³).

We found no evidence regarding the reliability of ES Reports; however the ARCP literature indicated a significant number may be poor quality.^{24, 26, 43} The Gold Guide refers to ES Reports as ‘structured’ and provides a general overview of the information that they should contain, although a template - which may improve the report’s reliability - is not provided. A non-systematic Google search revealed several templates in existence, and these varied considerably by medical specialty and region. It seems likely, therefore, that the quality and content of ES Reports – and therefore of the ARCP decisions based on them - vary considerably.

The way ARCP panels make decisions as a group can also affect reliability, as shown in the wider literature.⁴⁴⁻⁴⁶ A literature review on group decision-making⁴⁵ included recommendations for improving how Clinical Competency Committees (similar to ARCP panels) in the United States make decisions. The Accreditation Council for Graduate Medical Education has also produced an evidence-based guidebook⁴⁷ containing detailed guidance on enhancing the Clinical Competence Committees judgements. No such evidence-based guidance exists for the ARCP.

Fairness of the ARCP

The ARCP literature shows that on average trainees who qualified outside the UK, who are male, older, or from black and minority ethnic backgrounds are more likely to have an unsatisfactory ARCP outcome.^{21, 27, 35, 48} The reason for these differences may be multifactorial, although it is likely that they reflect the additional risk to achievement some groups experience during training.^{21, 49} There is also evidence that some trainees believe ARCP panels can be biased against minority ethnic and/or pregnant trainees.^{13, 49, 50}

Equality and diversity training – a requirement for all panel members – does not guarantee that panel decision-making will be fair. Indeed, Ahmed⁵¹ warns poor quality

training can conceal rather than guard against discrimination. Explicit discussion of equality and diversity during decision-making may remind assessors of their commitment to fairness as they make decisions.

Formative elements of the ARCP

Feedback given to trainees by the panel

National guidance states that trainees anticipated to receive an unsatisfactory ARCP outcome are required to attend the panel to discuss previous performance and plans to improve future performance ('feedforward'⁵²), and that such discussions should be separate from panel decision-making. The Foundation Programme Guidance also states that trainees with an *unanticipated* unsatisfactory outcome should have a feedforward meeting, that all trainees should have feedback about 'targeted learning, areas for improvement and/or areas of demonstrated excellence' (p.18)¹⁷ and the implication is for written feedback.

We found no estimate of what proportion of trainees attend a meeting with the panel, although the literature suggests that only some do.^{13, 53} For trainees who do attend, the research suggests that the separation of the ARCP decision-making process from feedback/feedforward may not always be clear.^{29, 54} Where feedback is provided, trainees can perceive it as unhelpful, negative or even confrontational.^{13, 24, 55, 56} Many trainees are also critical that the ARCP does not provide enough good quality feedback (e.g.^{21, 26, 52}).

The wider literature is clear that the benefits of feedback depend on the nature of the feedback, who it is delivered by, and how and when it is delivered.⁵⁷ The provision of constructive feedback and goals for improvement by ARCP panels is likely to encourage a culture of learning and development in which trainees aspire to excellence, and may therefore enhance patient safety. In the United States all trainees (residents/fellows) are required to

receive feedback after the Clinical Competency Committee, and guidance is issued on how feedback can be collated and delivered usefully.⁴⁷

In terms of trainees who are released from training following an ARCP, the Gold Guide does not mandate support, stating only they ‘may wish to seek further advice [...] about future career options.’(p.56)¹ By definition those required to leave training are likely to have performed very poorly. It makes sense educationally and for patient safety that ways be found to support them to develop their careers.

Preparing trainees for the ARCP

The literature suggests that when the ARCP was first introduced many trainees did not feel prepared^{25, 29} and it is not clear from the literature whether things have changed. Three studies reported on tools to support preparation by tracking achievement of competencies mapped to curricular requirements throughout the year.^{22, 58, 59} We also found an article describing how life coaching to address affective and attitudinal problems (rather than knowledge and skills problems) might help trainees with persistently poor ARCP performance.⁶⁰

We found examples of several LETBs who have introduced interim reviews, designed to support trainees in preparing for the ARCP. We found no formal evaluations of interim reviews, although one study reported trainee and trainer experiences of attending a pilot Annual Planning Meeting (APM) three months prior to the ARCP, at which feedforward was provided.⁵³ Trainees in the pilot found the APM encouraging, non-confrontational, and supportive, and liked that it did not rely on paperwork. Interestingly, trainers felt that only trainees in difficulty should have the meetings, but trainees who were progressing satisfactorily felt they gained from it, which supports the value of constructive, stretching feedback for all.

Impact on trainee motivation

The ARCP literature shows many trainees find the minimal competence aspect of the ARCP demotivating and discouraging of excellence.^{13, 24-26} The existence of several different categories of *unsatisfactory* ARCP outcomes (*cf.* a single ‘satisfactory’ category), and the manner in which these categories have been used, has contributed to a perception of the ARCP as negative, bureaucratic, and detrimental to learning.^{13, 21, 26}

There is much discussion in the medical education literature about minimal competence^{8-10, 12, 61, 62} and we cannot review it all here; however Eva and colleagues⁶³ argue compellingly that the concept is underpinned by the incorrect assumptions that a trainee who can perform a task well in one context can perform it equally well in all contexts, and that once competence at a task has been achieved and ‘ticked off’ a trainee no longer needs to work on it. This encourages trainees to learn just enough to achieve ‘sign-off’, and not to revisit a competence once it has been recorded as complete, which can hinder learning, result in poor performance, and endanger patients. Educators can also find it difficult to help just-passing trainees since they are rated as equivalent to trainees performing at an extremely high level.

Discussion

Summary of findings

We found relatively little published research assessing the validity of the ARCP. National ARCP guidance lacks detail, resulting in variable practice across locations and specialties, and threatening the validity and reliability of outcomes. Trainees and trainers have concerns ARCPs only identify very poorly performing trainees, which may arise partly from the ‘failure to fail’ phenomenon. The fact that feedback is not routinely provided to all trainees may leave those with specific or less serious performance issues unsupported and

demotivate high performers. Variability in the provision and quality of feedback from ARCP panels and when helping trainees to prepare for panels can negatively affect learning.

Strengths and limitations of the study

To our knowledge this is the first review of the ARCP, which is a fundamental aspect of postgraduate medical training in the UK, and has parallels in postgraduate medical training globally. Our study is strengthened by the systematic and inclusive search of the peer-reviewed and grey literature on ARCPs, by including information from national and local policy documents, and by comparing with the wider medical education literature on assessing competence. It was not possible to include all of the wider medical education literature on assessing competence, however we referenced reviews and highly cited papers where possible.

Implications for policy and practice

We have highlighted that increasing the standardisation of how ARCP decisions are made, how feedback is provided, and how trainees are prepared for the ARCP, are crucial to combat threats to the reliability of the ARCP. While national guidance cannot provide detailed information about specific curricular requirements, it could include information on: how panels should weight different pieces of information (which would need to be supported by further research); the expectation that assessments submitted in the portfolio are sampled across the curriculum and assessors; the need for the ES to be absent when panels are making decisions; the need to ensure trainee presence at a panel does not influence decision-making; and evidence-based ways to guard against bias arising from panel group decision-making. In addition, the quality of locally-generated tools for supporting decisions can be compared and standardised nationally.

Research is urgently needed to assess the predictive validity of the ARCP, in particular its ability to distinguish reliably between satisfactory and unsatisfactory performance and progress, to distinguish between the different levels of unsatisfactory performance, and to identify patient safety issues. This will require the collection and provision of good quality data to researchers, and the publication of findings. Initiatives such as the UK Medical Education Database⁶⁴ provide a mechanism for linking ARCP data with other outcomes and providing linked data to researchers, although the quality of such research will depend on the quality of the ARCP data. Fairness should also be considered as part of research into the validity of the ARCP, as is the case for other high-stakes assessments (e.g.^{65, 66}). As Tiffin and colleagues³⁵ point out, it is unlikely that the ARCP is ‘free from cultural influences and opportunities for assessor bias’.

The wider educational evidence suggests that ensuring all trainees receive constructive feedback to improve their learning and performance, including ‘stretching’ feedback for those performing well, will increase the educational value of the ARCP process and help motivate high performers. Providing all trainees with a pre-ARCP meeting with their ES and another person, possibly an ARCP panel member, to check progress and provide feedback can help ensure any problems are addressed early and will guard against ‘failure to fail’ by reducing the high-stakes nature of the final ARCP.

Conclusions

Assessment of trainees is necessary to ensure standards and protect patients. Epstein and Hundert⁶² state that assessment is ‘a statement of institutional values’ (p.231), and thus investment in developing the ARCP to the highest educational standards demonstrates the value placed on developing excellent doctors, and will help combat the current impression of the ARCP as a relatively ineffective, bureaucratic, box-ticking process. We suggest that

investment in undertaking high quality and continual evaluation of the ARCP is essential to ensure the validity, reliability, robustness, and defensibility of the ARCP and its role in postgraduate training.

Acknowledgements

We are grateful to Gloria Wong, a UCL medical student, for helping with the search for guidance documents. Thank you also to Catherine O’Keeffe for comments on earlier drafts of this paper.

Funding

This research was funded by Health Education England.

Competing interests

Katherine Woolf is funded by the National Institute for Health Research (NIHR) Career Development Fellowship (Grant Reference Number CDF-2017-10-008). The views expressed are those of the author and not necessarily those of the NIHR or the Department of Health and Social Care.

Author contributions

KW designed the study and conducted the literature review using search parameters and inclusion and exclusion criteria agreed by KW and MP in advance. KW analyzed the data and wrote the first draft of the original research report. MP redrafted the research report in collaboration with KW and RV, and all authors approved the final manuscript.

References

1. *A Reference Guide for Postgraduate Specialty Training in the UK: The Gold Guide*. 6th Edition ed.: Conference of Postgraduate Medical Deans of the United Kingdom, 2016.
2. Barrett A, Galvin R, Steinert Y, et al. A BEME (Best Evidence in Medical Education) systematic review of the use of workplace-based assessment in identifying and remediating poor performance among postgraduate medical trainees. *Systematic Reviews* 2015; 4: 65. journal article. DOI: 10.1186/s13643-015-0056-9.
3. Kogan JR, Holmboe ES and Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: A systematic review. *JAMA* 2009; 302: 1316-1326. DOI: 10.1001/jama.2009.1365.
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: 455-473. journal article. DOI: 10.1007/s10459-015-9614-0.
5. Miller A and Archer J. Impact of workplace based assessment on doctors' education and performance: a systematic review. *BMJ (Clinical research ed)* 2010; 341. DOI: 10.1136/bmj.c5064.
6. Moonen-van Loon JMW, Overeem K, Donkers HJLM, et al. Composite reliability of a workplace-based assessment toolbox for postgraduate medical education. *Advances in Health Sciences Education* 2013; 18: 1087-1102. journal article. DOI: 10.1007/s10459-013-9450-z.
7. Driessen E, Van Tartwijk J, Van Der Vleuten C, et al. Portfolios in medical education: why do they meet with mixed success? A systematic review. *Medical Education* 2007; 41: 1224-1233. DOI: doi:10.1111/j.1365-2923.2007.02944.x.
8. Holmboe ES, Sherbino J, Long DM, et al. The role of assessment in competency-based medical education. *Medical Teacher* 2010; 32: 676-682. DOI: 10.3109/0142159X.2010.500704.
9. ten Cate O. The false dichotomy of quality and quantity in the discourse around assessment in competency-based education. *Advances in Health Sciences Education* 2015; 20: 835-838. journal article. DOI: 10.1007/s10459-014-9527-3.
10. van der Vleuten CPM and Schuwirth LWT. Assessing professional competence: from methods to programmes. *Medical Education* 2005; 39: 309-317. DOI: 10.1111/j.1365-2929.2005.02094.x.
11. van der Vleuten CPM, Schuwirth LWT, Scheele F, et al. The assessment of professional competence: building blocks for theory development. *Best Practice & Research Clinical Obstetrics & Gynaecology* 2010; 24: 703-719. DOI: <https://doi.org/10.1016/j.bpobgyn.2010.04.001>.

12. Frank JR, Snell LS, Cate OT, et al. Competency-based medical education: theory to practice. *Medical Teacher* 2010; 32: 638-645. DOI: 10.3109/0142159X.2010.501190.
13. Viney R, Rich A, Needleman S, et al. The validity of the Annual Review of Competence Progression: a qualitative interview study of the perceptions of junior doctors and their trainers. *Journal of the Royal Society of Medicine* 2017; 110: 110-117. 2017/01/25. DOI: 10.1177/0141076817690713.
14. HEE to lead review to increase support for junior doctors career progression, <https://hee.nhs.uk/news-blogs-events/news/hee-lead-review-increase-support-junior-doctors-career-progression> (2016, accessed 21st June 2018).
15. Woolf K and Page M. *Academic support for the Assessment and Appraisal Workstream of Health Education England's review of the ARCP*. 2017. University College London.
16. Stobart G. Validity in formative assessment. In: Gardner J, (ed.). *Assessment and Learning*. 2 ed. London: SAGE Publications Ltd, 2012.
17. *Guide to the Foundation Annual Review of Competence Progression (ARCP) Process – 2017*. 2017. Foundation Programme.
18. *The Foundation Programme Reference Guide*. 2016. Foundation Programme.
19. *The GMC protocol for making revalidation recommendations: Guidance for Responsible Officers and Suitable Persons*. 2015. Manchester: General Medical Council.
20. *COPMeD guidance on making revalidation recommendations for doctors in postgraduate training*. 2017. Conference of Postgraduate Medical Deans of the United Kingdom.
21. Rothwell C. *A study to identify the factors that either facilitate or hinder medical specialty trainees in their Annual Review of Competence Progression (ARCP), with a focus on adverse ARCP outcomes* PhD Thesis, Durham University, 2017.
22. Ntatsaki E, Tugnet N, Nadesalingam K, et al. The pre-Annual Review of Competence Progression Checklist: Demystefying the Annual Review of Competence Preparation for Rheumatology. *Rheumatology* 2015; 54: 1.
23. McKee RF. The Intercollegiate Surgical Curriculum Programme (ISCP). *Surgery* 2008; 26: 411-416. Review. DOI: 10.1016/j.mpsur.2008.08.007.
24. Eynon-Lewis A and Price M. Reviewing the ARCP process: experiences of users in one English deanery. *BMJ Careers* 2012.
25. Peiris L and Cresswell B. How to succeed in the Annual Reviews of Competency Progression (ARCP). *Surgery (United Kingdom)* 2012; 30: 455-458. Article. DOI: 10.1016/j.mpsur.2012.06.004.

26. Dormandy L and Laycock K. Triumph of process over practice: changes to assessment of physicians. *BMJ Careers* 2015.
27. Bedward J, Davison I, Burke S, et al. *Evaluation of the RCGP GP Training Curriculum*. 2011. University of Birmingham, University of Warwick.
28. Bodgener S, Denney M and Howard J. Consistency and reliability of judgements by assessors of case based discussions in general practice specialty training programmes in the United Kingdom. *Education for Primary Care* 2017; 28: 45-49. DOI: 10.1080/14739879.2016.1229579.
29. Goodyear H, Wall D and Bindal T. Annual review of competence: Trainees' perspective. *Clinical Teacher* 2013; 10: 394-398. Article. DOI: 10.1111/tct.12040.
30. van der Vleuten CPM. A Programmatic Approach to Assessment. *Medical Science Educator* 2016; 26: 9-10. journal article. DOI: 10.1007/s40670-016-0343-7.
31. Cho SP, Parry D and Wade W. Lessons learnt from a pilot of assessment for learning. *Clinical Medicine* 2014; 14: 577-584. DOI: 10.7861/clinmedicine.14-6-577.
32. van der Vleuten CPM, Schuwirth LWT, Driessen EW, et al. A model for programmatic assessment fit for purpose. *Medical Teacher* 2012; 34: 205-214. DOI: 10.3109/0142159X.2012.652239.
33. van der Vleuten CPM, Schuwirth LWT, Driessen EW, et al. Twelve Tips for programmatic assessment. *Medical Teacher* 2015; 37: 641-646. DOI: 10.3109/0142159X.2014.973388.
34. Ludka-Stempien K. *Predictive validity of the examination for the Membership of the Royal Colleges of Physicians of the United Kingdom*. University College London, London, 2015.
35. Tiffin PA, Illing J, Kasim AS, et al. 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 (Clinical research ed)* 2014; 348: g2622. 2014/04/20. DOI: 10.1136/bmj.g2622.
36. Gale TCE, Roberts MJ, Sice PJ, et al. Predictive validity of a selection centre testing non-technical skills for recruitment to training in anaesthesia. *BJA: British Journal of Anaesthesia* 2010; 105: 603-609. DOI: 10.1093/bja/aeq228.
37. Pashayan N, Gray S, Duff C, et al. Evaluation of recruitment and selection for specialty training in public health: interim results of a prospective cohort study to measure the predictive validity of the selection process. *Journal of public health (Oxford, England)* 2016; 38: e194-200. 2015/08/13. DOI: 10.1093/pubmed/fdv102.
38. Davison I, McManus I and Taylor C. *Evaluation of GP Specialty Selection*. 2016. University of Birmingham, University College London, University of Warwick.

39. Wakeford R, Ludka K, I; M, et al. Poor performance on postgraduate medical examinations predicts increased likelihood of regulatory sanctions in UK-registered doctors *Annual European Conference on Assessment in Medical Education*. Egmond Ann Zee, Netherlands: European Board of Medical Assessors, 2017.
40. Yepes-Rios M, Dudek N, Duboyce R, et al. The failure to fail underperforming trainees in health professions education: A BEME systematic review: BEME Guide No. 42. *Medical Teacher* 2016; 38: 1092-1099. DOI: 10.1080/0142159X.2016.1215414.
41. *Generic professional capabilities: guidance on implementation for colleges and faculties*. 2017. Manchester: Academy of Medical Royal Colleges and the General Medical Council.
42. Driessen E and Scheele F. What is wrong with assessment in postgraduate training? Lessons from clinical practice and educational research. *Medical Teacher* 2013; 35: 569-574. DOI: 10.3109/0142159X.2013.798403.
43. Edwards J and Petra H. The effects of external quality management on workplace-based assessment. *Education for Primary Care* 2013; 24: 105-110. Article.
44. Hauer KE, Chesluk B, Iobst W, et al. Reviewing Residents' Competence: A Qualitative Study of the Role of Clinical Competency Committees in Performance Assessment. *Academic Medicine* 2015; 90: 1084-1092. DOI: 10.1097/acm.0000000000000736.
45. Hauer KE, ten Cate O, Boscardin CK, et al. Ensuring Resident Competence: A Narrative Review of the Literature on Group Decision Making to Inform the Work of Clinical Competency Committees. *Journal of Graduate Medical Education* 2016; 8: 8. DOI: 10.4300/JGME-D-15-00144.1.
46. Chahine S, Cristancho S, Padgett J, et al. How do small groups make decisions? *Perspectives on Medical Education* 2017; 6: 192-198. journal article. DOI: 10.1007/s40037-017-0357-x.
47. Andolsek K, Padmore J, Hauser KE, et al. *Clinical Competency Committees A Guidebook for Programs*. Accreditation Council for Graduate Medical Education (ACGME), 2015.
48. Pyne Y and Ben-Shlomo Y. Older doctors and progression through specialty training in the UK: a cohort analysis of General Medical Council data. *BMJ Open* 2015; 5: e005658. 2015/02/05. DOI: 10.1136/bmjopen-2014-005658.
49. Woolf K, Rich A, Viney R, et al. Perceived causes of differential attainment in UK postgraduate medical training: a national qualitative study. *BMJ Open* 2016; 6. DOI: 10.1136/bmjopen-2016-013429.
50. Rich A, Viney R, Needleman S, et al. 'You can't be a person and a doctor': the work-life balance of doctors in training—a qualitative study. *BMJ Open* 2016; 6: e013897. DOI: 10.1136/bmjopen-2016-013897.

51. Ahmed S. 'You end up doing the document rather than doing the doing': Diversity, race equality and the politics of documentation. *Ethnic and Racial Studies* 2007; 30: 590-609. DOI: 10.1080/01419870701356015.
52. Rees CE, Cleland JA, Dennis A, et al. Supervised learning events in the Foundation Programme: a UK-wide narrative interview study. *BMJ Open* 2014; 4. DOI: 10.1136/bmjopen-2014-005980.
53. Bindal T, Wall D and Goodyear H. Annual planning meetings: views and perceptions. *The clinical teacher* 2014; 11: 524-530. DOI: 10.1111/tct.12193.
54. Vance G, Williamson A, Frearson R, et al. Evaluation of an established learning portfolio. *The clinical teacher* 2013; 10: 21-26. 2013/01/09. DOI: 10.1111/j.1743-498X.2012.00599.x.
55. Vasudev A, Vasudev K and Thakkar P. Trainees' perception of the Annual Review Of Competence Progression: 2-Year survey. *Psychiatrist* 2010; 34: 396-399. Review. DOI: 10.1192/pb.bp.109.028522.
56. Vasudev A, Thakkar P and K V. The 1st Annual Review of Competence Progression, a new way of assessing trainee doctors: Trainees' perception. *Medical Teacher* 2010.
57. Brennan N, Bryce M, Pearson M, et al. Towards an understanding of how appraisal of doctors produces its effects: a realist review. *Medical Education* 2017; 51: 1002-1013. DOI: 10.1111/medu.13348.
58. Ap Dafydd D, Williamson R, Blunt P, et al. Development of training-related health care software by a team of clinical educators: their experience, from conception to piloting. *Advances in medical education and practice* 2016; 7: 635-640. 2016/11/18. DOI: 10.2147/amep.s108426.
59. Wentworth L, Wardle K, Ruddlesdin J, et al. Members' Presentations Abstract: Competency Mapping in Quality Management of Geriatric Medicine training – A Survey of Trainees in the North-Western Deanery. *Medical Education* 2011; 45: 1-85. DOI: 10.1111/j.1365-2923.2011.04093.x.
60. Gale AC, Gilbert J and Watson A. Life coaching to manage trainee underperformance. *Medical Education* 2014; 48: 539-539. DOI: 10.1111/medu.12436.
61. Bok HG, Teunissen PW, Favier RP, et al. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Medical Education* 2013; 13: 123. journal article. DOI: 10.1186/1472-6920-13-123.
62. Epstein RM and Hundert EM. Defining and assessing professional competence. *JAMA* 2002; 287: 226-235. DOI: 10.1001/jama.287.2.226.
63. Eva KW, Bordage G, Campbell C, et al. Towards a program of assessment for health professionals: from training into practice. *Advances in Health Sciences Education* 2016; 21: 897-913. journal article. DOI: 10.1007/s10459-015-9653-6.

64. Dowell J, Cleland J, Fitzpatrick S, et al. The UK medical education database (UKMED) what is it? Why and how might you use it? *BMC Medical Education* 2018; 18: 6. journal article. DOI: 10.1186/s12909-017-1115-9.

65. Denney ML, Freeman A and Wakeford R. MRCGP CSA: are the examiners biased, favouring their own by sex, ethnicity, and degree source? *British Journal of General Practice* 2013; 63: e718-e725. DOI: 10.3399/bjgp13X674396.

66. McManus IC, Elder AT and Dacre J. Investigating possible ethnicity and sex bias in clinical examiners: an analysis of data from the MRCP(UK) PACES and nPACES examinations. *BMC Medical Education* 2013; 13: 103. journal article. DOI: 10.1186/1472-6920-13-103.

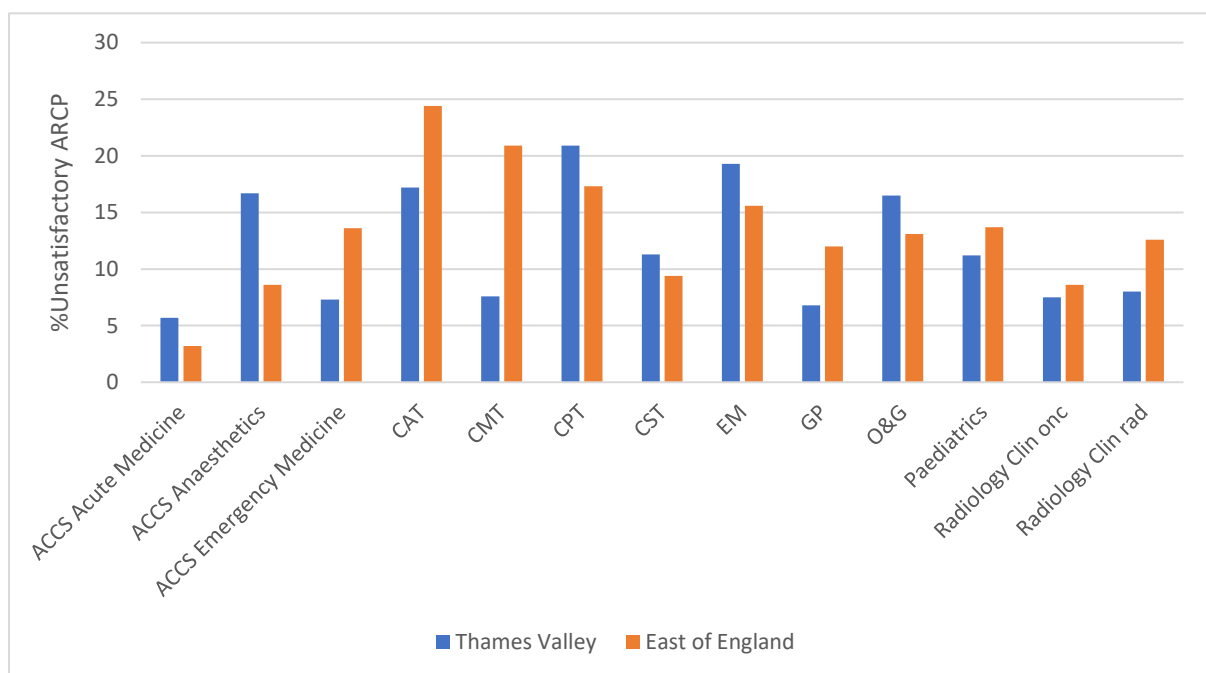


Figure 1: Variability in the percentage of ‘unsatisfactory’ ARCP outcomes awarded to trainees from 2010 to 2016 across 13 specialties in two Health Education England Local Education and Training Boards: Thames Valley and East of England. ACCS= Acute Care Common Stem; CAT=Core Anaesthetics Training; CMT=Core Medical Training; CPT=Core Psychiatry Training; CST=Core Surgical Training; EM=Emergency Medicine; GP=General Practice; O&G = Obstetrics & Gynaecology; Clin onc= Clinical

Oncology; Clin rad=Clinical Radiology. Data from <http://www.gmc-uk.org/education/14105.asp>

Supplementary Table 1: Local and/or specialty-specific guidance and decision aids included in the review.

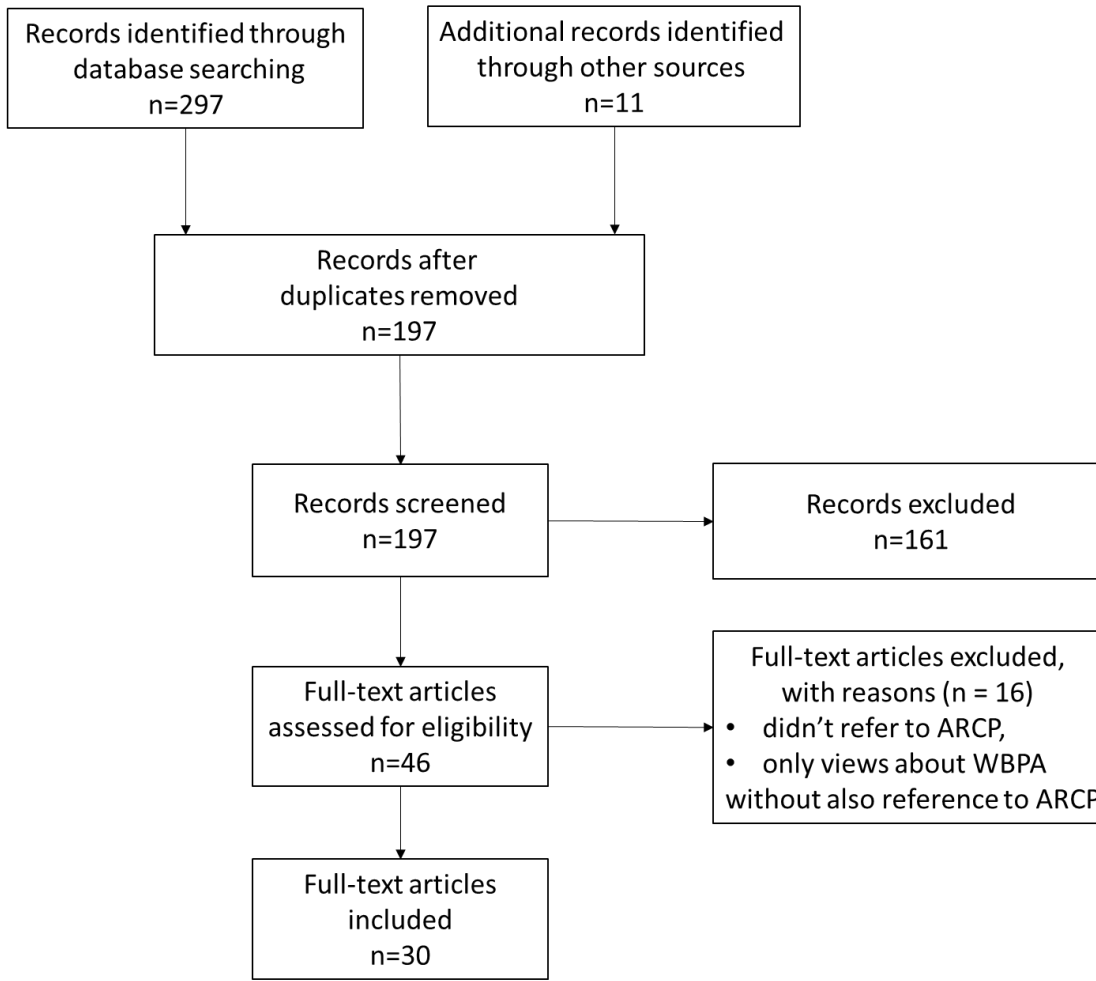
Body	Year	Title	Weblink
Joint Royal Colleges of Physicians Training Board	2014	General Internal Medicine (GIM) ARCP Decision Aid Revised 2014	https://www.jrcptb.org.uk/sites/default/files/2009%20GIM%20ARCP%20Decision%20Aid%20%28revised%2017.1.14%29.pdf
Joint Royal Colleges of Physicians Training Board	2015	2010 Cardiology (amendments 2015) ARCP Decision Aid	https://www.jrcptb.org.uk/sites/default/files/2010%20Cardiology%20ARCP%20Decision%20Aid%20%28revised%202015%29.pdf
Joint Royal Colleges of Physicians Training Board	August 2014	2010 Paediatric Cardiology ARCP Decision Aid	https://www.jrcptb.org.uk/sites/default/files/2010%20Paediatric%20Cardiology%20ARCP%20Decision%20Aid%20(revised%20January%202016).pdf
Joint Royal Colleges of Physicians Training Board	2014	Recommendations for specialty trainee assessment and review: Incorporating	https://www.jrcptb.org.uk/sites/default/files/April%202014%20Recommendations%20for%20specialty%20trainee%20assessment%20and%20review.pdf

		lessons learnt from the workplace-based assessment pilot	
Joint Committee on Surgical Training	nd	Certification (CCT/CESR (CP)) checklist for completion at by Trainee and review by LM BEFORE the final ARCP Otolaryngology Specialty Advisory Committee (SAC)	https://www.jcst.org/quality-assurance/certification-guidelines-and-checklists/
Bradford General Practice Vocational Training Scheme	nd	ARCP webpages	http://www.bradfordvts.co.uk/educational-supervision/arcp/
Royal College of Pathologists	nd	Histopathology ARCP Decision Aid – Stage D (ST5).	https://www.rcpath.org/asset/58F0F8B3-9216-4CE8-B06C5EA601264583.307D505A-41A4-4D34-AACAC436A607E102/

Health Education England Thames Valley	2015	The Educational Supervisor's Structured Report (Summative Assessment) for the ARCP.	http://www.oxforddeanery.nhs.uk/pdf/ES%20final%20%20report%20instructions2015.pdf
Health Education England West Midlands		ARCP Information Sheet	https://www.westmidlandsdeanery.nhs.uk/LinkClick.aspx?fileticket=RxlruKBkkW8%3D&tabid=536&portalid=0&mid=1472
Royal College of Psychiatrists	April 2008	ARCP Guidance 2008	http://www.rcpsych.ac.uk/docs/ARCP%20Guidance%20April%202008%20Final.doc
Royal College of Psychiatrists	May 2014	ARCP for Less Than Full-time & Academic Trainees	https://www.rcpsych.ac.uk/pdf/ARCP_Guidance_for_LTF_T_and_Academics_May2014.pdf

Supplementary Table 2: Summary of final full-text studies and reports. Main topic addressed in the reports: 1=Statistical relationships with other variables; 2= Demographic differences in outcomes; 3= Stakeholder perceptions; 4=Evaluation or description of process; 5= Preparation and feedback.

Authors	Year	Report theme	Source	Report type
Tiffin <i>et al</i> ;	2014	1	BMJ	Research paper
Pashayan <i>et al</i> .	2016	1	Journal of Public Health	Research paper
Ludka-Stempien	2015	1	University College London	PhD thesis
Goodyear <i>et al</i>	2013	1	Clinical Teacher	Research paper
Gale <i>et al</i>	2010	1	British Journal of Anaesthesia	Research paper
Davison <i>et al</i>	2016	1	Health Education England	Research report
Burnand <i>et al</i>	2014	1	R C S Bull	Research paper
Bodgener <i>et al</i>	2017	1	Education for Primary Care	Research paper
Bedward <i>et al</i>	2011	1,2	University of Birmingham	Research report
Rangan <i>et al</i>	2017	2	BMJ Open	Research paper
Pyne, Ben-Shlomo	2015	2	BMJ Open	Research paper
McKee	2008	2	Surgery	Description
General Medical Council	2016	2	General Medical Council	Research paper
Rothwell	2017	2,3	University of Durham	PhD thesis
Peiris, Cresswell	2012	2,3	Surgery	Opinion
Viney <i>et al</i> .	2017	3	J R Soc Med	Research paper
Vasudev <i>et al</i>	201	3	Psychiatrist	Research paper
Vasudev <i>et al</i>	2010	3	Medical Teacher	Research paper
Vance <i>et al</i>	2013	3	Clinical Teacher	Research paper
Goodyear <i>et al</i>	2012	3	Medical Teacher	Research paper
Eynon-Lewis, Price	2012	3	BMJ Careers	Careers article
Dormandy, Laycock	2016	3	BMJ Careers	Careers article
Cho <i>et al</i>	2014	3	Clinical Medicine	Research paper
Edwards, Petra	2013	4	Education for Primary Care	Research report
Black	2013	4	Clinical Medicine	Opinion
Wentworth <i>et al</i>	2011	5	Medical Education	Abstract
Ntatsaki <i>et al</i>	2015	5	Rheumatology	Poster
Gale <i>et al</i>	2014	5	Medical Education	Brief report
Bindal <i>et al</i>	2014	5	The Clinical Teacher	Research paper
Ap Dafydd <i>et al</i>	2016	5	Adv Med Ed and Prac	Description



Supplementary Figure 1: PRISMA Flow diagram for the ARCP literature search