



7-1-2012

UK and US Physician Assistant Student Performance on an International Test of Core Biomedical Science Knowledge

Scott Arbet
Arbet Consulting Inc

Jimmy Parle
University of Birmingham

Janet J. Lathrop
National Commission on the Certification of Physician Assistants

Follow this and additional works at: <https://nsuworks.nova.edu/ijahsp>



Part of the [Medicine and Health Sciences Commons](#)

Recommended Citation

Arbet S, Parle J, Lathrop JJ. UK and US Physician Assistant Student Performance on an International Test of Core Biomedical Science Knowledge. *The Internet Journal of Allied Health Sciences and Practice*. 2012 Jul 01;10(3), Article 5.

This Manuscript is brought to you for free and open access by the College of Health Care Sciences at NSUWorks. It has been accepted for inclusion in *Internet Journal of Allied Health Sciences and Practice* by an authorized editor of NSUWorks. For more information, please contact nsuworks@nova.edu.

UK and US Physician Assistant Student Performance on an International Test of Core Biomedical Science Knowledge

Abstract

Physician assistants (PAs) have provided cost effective, high quality care in the United States for more than 40 years. There is a growing international interest in utilizing PAs for a variety of pressing health care needs. This pilot study compared the performance of PA students trained in the United States to those trained in the United Kingdom on a core biomedical science knowledge examination using multiple choice questions developed by the National Commission on Certification of Physician Assistants (NCCPA). The study found that despite differing educational models and health systems, the students performed similarly. While rigorous statistical analyses were not possible given the small sample sizes, the study provides a promising indication that there is an international common core of biomedical science knowledge. Repeated studies and the expansion of the pilot to other countries will provide more generalizability and statistical support to establish whether there is an assessable, global core of PA biomedical science knowledge that could become one component of locally determined national standards for PAs.

UK and US Physician Assistant Student Performance on an International Test of Core Biomedical Science Knowledge

Scott Arbet, PhD¹
Jim Parle, FRCGP, MD²
Janet J Lathrop, MBA³

1. President, Arbet Consulting Inc, Malvern, Pennsylvania
2. Professor, Primary Care Clinical Sciences, University of Birmingham, Edgbaston, Birmingham
3. President and CEO, National Commission on the Certification of Physician Assistants, Johns Creek, Georgia

United States
United Kingdom

CITATION: Arbet S, Parle J, Lathrop J. UK and US Physician Assistant Student Performance on an International Test of Core Biomedical Science Knowledge. *The Internet Journal of Allied Health Sciences and Practice*. July 2012. Volume 10 Number 3.

ABSTRACT

Physician assistants (PAs) have provided cost effective, high quality care in the United States for more than 40 years. There is a growing international interest in utilizing PAs for a variety of pressing health care needs. This pilot study compared the performance of PA students trained in the United States to those trained in the United Kingdom on a core biomedical science knowledge examination using multiple choice questions developed by the National Commission on Certification of Physician Assistants (NCCPA). The study found that despite differing educational models and health systems, the students performed similarly. While rigorous statistical analyses were not possible given the small sample sizes, the study provides a promising indication that there is an international common core of biomedical science knowledge. Repeated studies and the expansion of the pilot to other countries will provide more generalizability and statistical support to establish whether there is an assessable, global core of PA biomedical science knowledge that could become one component of locally determined national standards for PAs.

BACKGROUND

For more than 40 years, the United States has utilized physician assistants (PAs) as part of the health care team to fill evolving health care needs, to combat provider shortages, and to foster continuity of care. This still growing part of the US medical workforce includes more than 80,000 certified PAs, which represents considerable growth from the three students who completed the first PA program at Duke University in October 1967. Trained in the medical model, PAs practice medicine with the general supervision of a physician. In the US, PAs must graduate from an educational program accredited by the Accreditation Review Commission on Education for the Physician Assistant (ARC-PA), pass a national certifying exam within six attempts or six years of their program completion, and then, in most cases, attain licensure (or registration) from a state medical authority to practice. Specific licensure requirements vary by state, but all 50 states and US territories rely on national certification as one of the criteria for initial licensure.

Since 1975, the National Commission on Certification of Physician Assistants (NCCPA) has been responsible for establishing initial (and recertifying) examinations for PAs to test the knowledge and skills needed for practice. The examination is based on a practice analysis, also called a role delineation study, which is a systematic plan to study a profession, and a critical step in

developing a psychometrically sound and reliable credentialing examination. A practice analysis forms the link between examination content and real world clinical performance by identifying the knowledge and skills PAs use in their daily practice, as well as the frequency and criticality of activities related to that knowledge and skill. The result is an examination content blueprint that is matched to what clinically active PAs are presently doing in practice across the body's organ systems (e.g., cardiovascular system, pulmonary system, GI, etc.).¹

NCCPA scores its computer-based, 300 multiple choice question (MCQ) examination using a criterion-referenced standard. The NCCPA Board of Directors sets the standard based on recommendations from standard setting committees that use the Hofstee and modified Angoff methodologies, which are legally defensible approaches. DeChamplain describes the Hofstee method as a holistic method that focuses on overall acceptability of percent-correct passing scores, as well as the percentages of examinees who might fail the examination.² In comparison, DeChamplain also explains the modified Angoff methodology requires independent review of each item, first without statistical performance data, and then a second rating with statistical information, reflection on the definition of minimum competence, and an estimation of the probability that a borderline, minimally competent candidate would correctly respond to that specific item.² (See additional discussion on the Angoff methodology below). NCCPA reevaluates the job analysis and the passing standard approximately every two to three years.

In an effort to address economic pressures, reduced hours worked by physicians in training, and concerns over continuity of care, there has been international interest in the adoption of the PA concept over the past 10 years. Specifically, interest in this concept has emerged in Australia, Canada, Ghana, Germany, Ireland, the Netherlands, New Zealand, Scotland, South Africa, and the United Kingdom. In 2010, the NCCPA invited academicians, clinicians, and others from nine countries to an informal meeting where NCCPA shared lessons learned during its 35 years of experience as the national PA certification body. Interest in exploring whether there was a core global knowledge base common to all PAs, and whether that knowledge base is assessable, emerged over the two-day discussion. However, delegates agreed that different countries would also wish to test for local clinical knowledge and ethical, sociological, and legal issues (e.g. attitudes regarding consent, or the law on termination of pregnancy). The UK PA programs were represented by the Chair of the UKIUBPAE (United Kingdom and Ireland Universities Board for PA Education, which has a role similar to the Physician Assistant Education Association) and the Chair of the exam sub-committee of the UKIUBPAE (this committee has a role similar to NCCPA), among others. Based on the discussion, the idea for this pilot study to assess the possibility of sharing items and comparing educational models and systems between the UK and the US emerged.

NATIONAL ASSESSMENT FOR PAs IN THE UNITED KINGDOM

Although there is not yet a legal requirement for UK-trained PAs to take a national assessment, PA programs and their universities all agree that conferment of a post graduate diploma in Physician Assistant Studies was contingent not only on passing the university's internal assessments, but also on the success in a national exam. This decision is in keeping with the recommendations of the Competence and Curriculum Framework (2006) produced by the UK government.³ The national assessment, which has been in place for four years, consists of a 200 MCQ examination and a 16-station OSCE (Objective Structured Clinical Examination) in which students move around a circuit of 8-minute clinical "stations" with 2-minutes preparation between each station. This OSCE is intended to bear some resemblance to short clinical encounters or ward rounds. Thus, the PA may be challenged to interact with a patient (e.g. taking a history or giving information on test results), to perform a focused clinical examination on a simulated patient (e.g. who is presenting with a history of weight loss and tremor suggestive of hyperthyroidism), or to suggest a management strategy (e.g. explaining to a supervising physician what is the appropriate next management step for a patient with jaundice). Students have to pass both assessments independently achieving 50% correct in the MCQ (see discussion below about standard setting) and passing 12 of the 16 OSCE stations. Students may have up to three attempts at each assessment.

In the UK, the passing standard for the MCQ test is established by using the Angoff method in which, as described above, content experts determine how difficult a question is for borderline candidates taking the exam at this stage in their career (since exactly the same question on, for example, insulin dosage, may be easy for an experienced diabetologist and difficult for a 3rd year medical student). If the question is relatively easy, the Angoff score (the percentage of borderline candidates who would get the answer correct) will be high, and vice versa. Experts generally complete this exercise individually and then as a group. The results of the Angoff exercise can then be mathematically translated to a percent correct score resulting in the "official" pass mark/standard. Once the content-based standard is set, passing the examination is based solely on the amount of content examinees answered correctly. Some people would argue that the pass mark/standard should indeed be on the order of 70% correct to pass. The relative difficulty of a test impacts the Angoff passing standard to a point; however, content experts are rarely able to accurately identify easy and hard test items, and are generally not apt to change their initial item ratings once they have seen the actual item performance data. It is very common for Angoff based standard setting ratings to yield results near

70% correct. This is also due to the fact that, for many of us in school, 70% on classroom tests represents "just passing" the class.

For OSCEs, perhaps even more than for MCQs, setting passing standards is the subject of considerable debate, and most methods depend on having large numbers of candidates taking the test at one time. For the UK national PA exam, global scoring descriptions are used that depend on the holistic judgments of expert clinicians (expert both in the subject and in the reasonable expectations of candidates at this level, as is the case for Angoff scoring) to decide if a candidate has passed or failed a station. Therefore, a student will usually be seen by 16 different examiners, each making independent pass/fail decisions.

METHODOLOGY

A first step in determining the viability of this pilot study, to compare exam performance of US and UK-trained PAs on a core knowledge component, was to ascertain whether NCCPA-developed items were appropriately transferrable to students educated in the UK. NCCPA supplied the Chair of the exam subcommittee of the UK board with over 500 MCQ items to determine the validity and transferability of the items. UK experts performed an initial review of the items and determined that approximately 85-90% of the items provided by NCCPA would be applicable, with only minimal revision to Anglicize the items and/or perform appropriate unit conversions (e.g., pounds to kilograms).

The items were then reviewed in depth by a group of 10 clinicians, including UK and US trained PAs. Their first goal was to verify the validity of the items for UK PAs in order to exclude those not appropriate (sometimes because of diseases not seen in the UK, most often because of differences in management or available medications). Secondly, items were excluded if they did not include five response choices. The group then established a standard for the questions using the Angoff methodology. From the accepted NCCPA items, plus additional UK-specific items developed by the UK exam subcommittee, two overlapping exam forms with 200 MCQ items were developed in accordance with the content blueprint included in the Competence and Curriculum Framework, which specifies the nationally-agreed upon content for UK-trained PAs. Recognizing that an MCQ can only sample a clinical area, a minimum number of questions per area was also required.

The forms were administered to UK PA students at or very close to the end of their program on two separate occasions in the fall of 2010. As is normal for MCQ examinations in the UK, a post-hoc review of question performance was conducted by three members of the examination subcommittee. Questions were excluded from the assessment if they performed exceptionally poorly. For example, if less than 30% of students correctly answered a question AND the correlation was low (in essence, if the students scoring best overall got the question incorrect), the item was considered unacceptable and therefore removed. The remaining items were then scored, the passing mark/standard applied, and the results were reported.

RESULTS

A total of 170 NCCPA-derived questions were selected for the first paper of 2010, and a total of 46 candidates took the assessment on two separate occasions. For the UK students, a range of raw scores from 48 to 81 percent correct were recorded with an overall average percent correct for the exam of 63.66% and an overall standard deviation of ± 17.83 . These results would seem to indicate a fairly heterogeneous group; however, given the small sample size, caution should be exercised in drawing this conclusion due to the possibility that outliers can have a more pronounced effect. By comparison, NCCPA has item performance data on these items for a significant number of first-time test takers (although the number of first time takers who have seen each item varies) with an overall mean percent correct of 75.07% and an overall standard deviation of ± 12.45 . The smaller standard deviation would indicate a more homogeneous group of students, which is likely attributable to long-established programs and curricula.

The performance results suggest that UK first-time takers represent a more heterogeneous group than current US first-time takers, which could be attributable to myriad reasons, including item differences with respect to content coverage, the vignette style of the questions, and the diversity of academic and clinical experiences for the UK students who took the examination (e.g., classical model: didactic followed by clinical; integrated model: problem based learning, simulation, early clinical contact, interweaving of theory and practice; or a mixed model).

Given the small sample size of the UK population and relatively small number of items for which there was UK performance data (for US takers there is performance data on thousands of candidates), few additional statistical analyses were deemed prudent beyond comparing UK student performance to the US first-time taker national averages as described. Recognizing, however, that there is variability in every population, it was agreed that additional descriptive data could be provided because NCCPA uses an automated test assembly system whereby test forms are built to be "nominally parallel" with respect to the content blueprint, item difficulty, the standard deviation of the difficulty level of items, the number of seconds on average to answer each question,

and other variables. These test forms are pre-built and pre-equated, making them essentially mirror images of each other. Each test form is interchangeable with virtually no practical differences. Therefore, one can make the "leap" that the items seen by the UK students, which had similar distributional characteristics and came from the same item pool as those seen by the first-time test takers from US programs, are similar enough to make valid comparisons.

With the US first-time taker national average for percent correct at approximately 75% and the UK average for the pilot at 64%, it was suggested that an interesting relative comparison may be to identify what the national average percent correct is for the lowest performing (in terms of first-time takers' PANCE performance) 10% of US PA programs in 2008, 2009, and 2010. This group represented 21 programs out of the current number of 156 accredited programs as of March 2011, and the included programs changed slightly over the three-year period. (Programs testing fewer than 10 students were excluded from consideration.) The average weighted percent correct for these programs was slightly over 68%. The percentage was calculated based on item performance data from thousands of items and across the test forms administered for each of the three years.

This finding illustrates that the performance of the UK students, who demonstrated an average percentage correct of slightly fewer than 64% is more similar to the performance of the lowest 10% of US programs. This finding reinforces the consistency of the scoring for test items based on both a sample of items that meet very similar characteristics as we have used for multiple different test forms over three years. Additional research and data collection may better illuminate the potential for advancements for the UK and other international programs.

LIMITATIONS & IMPLICATIONS FOR FUTURE RESEARCH

As briefly noted, two limitations for this study were the small number of UK students who took the national assessment, and the varying academic and clinical experiences of those students. This would suggest that repeated analyses, as additional cohorts take successive examinations, might provide more insight into the performance of UK PAs on similar items as were previously administered to US first-time test takers. At present, the small sample size emphasizes the variability in the data represented by the students who did very well or very poorly, and undermines the generalizability of the results. Another limitation of the study may be that consideration was not given to the diverse demographics of the students. For example, the average age of a US first-time exam taker is approximately 26 with approximately two years of healthcare background, while the average age for a UK exam taker was 32 with seemingly a bit more healthcare experience (albeit this information was provided anecdotally) and a longer gap since their own first degree experience.

Although potentially comparable, the data may indicate that UK students are slightly less able on average and perform on the lower end of the spectrum compared to US students. Consequently, a follow-up study may be indicated to compare the UK students' performance with historical US performance from when the profession/certifying examination was first developing and/or when new programs graduate their first few cohorts of students. Such a study might better equate the sample sizes and better reflect environmental factors such as experience of PA faculty, variability in curriculum, etc., that may be current challenges of the UK programs. Another possible comparison that could shed light on the numerous variables that could account for the performance differences would be a comparison of the admissions criteria for UK and US programs. This could suggest that the students themselves are a bit less able, or it might suggest that the instruction in the UK does not match the assessment (as measured by the NCCPA test questions) or variability in the quality of instruction in UK and US PA programs.

In general, the implications of the pilot study are very encouraging. The results demonstrate that the UK PA profession is progressing, particularly given the relative newness of the profession and the national exam (4 years) compared to the nearly 40 years of experience in the US. Further, the pilot study's results suggest validation for the use of NCCPA-developed items by students in the UK (who completed differing educational programs and will practice in differing health systems). This validation came from both the content experts' review and statistical item validation. The implication of this finding is that in addition to continuing the UK pilot, it may be appropriate to expand the pilot to include other countries with PA educational programs. Further, it may also be appropriate to determine whether items developed in the UK or elsewhere could be used on NCCPA examinations. The exciting possibility potentially inherent in the sharing of exam items may be to identify a global core biomedical science knowledge base for PAs and then to build a shared item bank to support that knowledge base.

CONCLUSION

The encouraging results of the present study, based on the comparative performance of US and UK-trained students on NCCPA-developed items, suggest that it may be feasible to develop a pool of test items from which examinations could be constructed to assess PAs in different countries with different educational and health systems. The next most appropriate step is to replicate this study with future UK cohorts and expand the pilot to include other countries that have developed PA educational programs and a mechanism for administering a secure national exam to all students. The most likely countries, which may be

ready in the short-term to participate in future studies, would be Australia and Canada. Long-term, the possibility for increased cooperation to develop global standards of core biomedical science knowledge for PAs is most impressive and exciting, as similar standards are not known to be available for any other health professions. Nonetheless, the foundation for success – even with the potential for international collaboration on a core examination component – remains having confidence in the country-specific development of PAs, modeling on, but not wholesale adoption of, US systems, and supplementing a core global knowledge base with appropriate local assessments. The end result, the continued sharing of the PA concept, is positioned to help us achieve positive results for the larger and more pressing challenge: to supplement the clinical workforce globally.

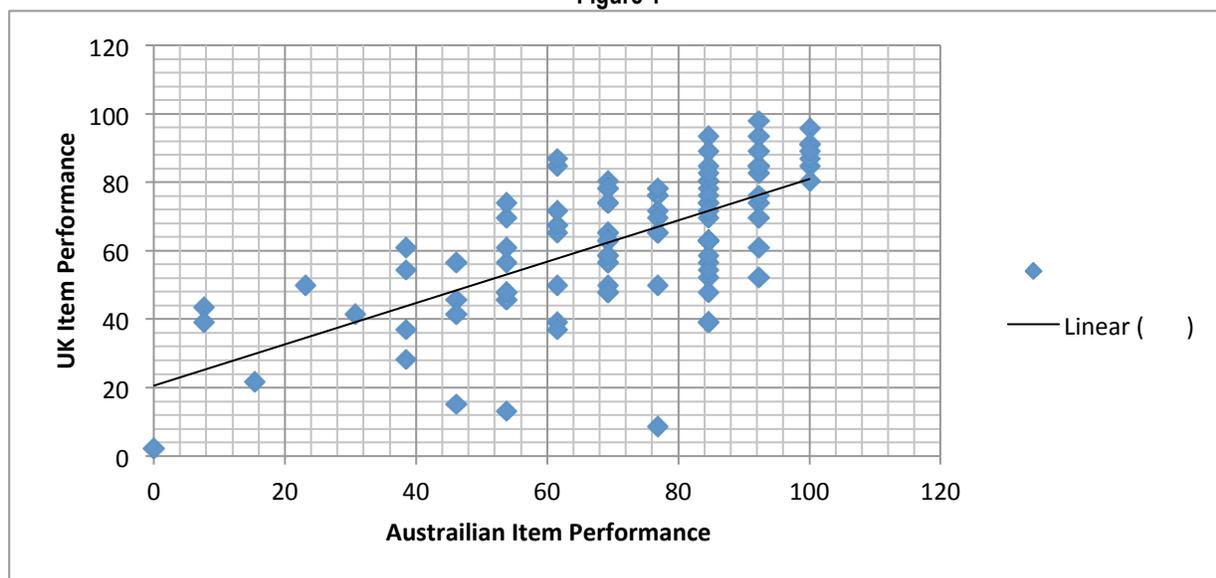
ADDENDUM

Since this investigation (Fall 2010), in the summer of 2011 a subset of 100 of the same 170 NCCPA-derived questions were administered to 13 candidates in Australia (It should be noted that there was an administrative error by Australia in the development of the test resulting in one of the 100 items being invalidated, thus only 99 items remained). As a reminder, based on the 170 items scored for the UK students, a range of raw scores from 48% to 81% correct were recorded with an overall average percent correct for the exam of 63.66% and a standard deviation of ± 17.83 . The average percentage correct for UK students on the subset of 99 items in common with Australia was 63.26.

Based on the common 99 items, the Australian candidates had an average total test score of 70.71 with a standard deviation of ± 5.44 (total scores ranged from a total percent correct score of 63.36 to 80.81). These results seem to indicate similar general performance, noting however, that the Australian students attained a higher mean score on the 99 common items of approximately 8 percentage points. Apparently, the UK group of candidates is more heterogeneous; however, given the small sample size, as in the case of the UK pilot study, caution should be exercised in drawing any ability conclusions as a result of the possibility of outliers when, for example, comparisons are made with NCCPA performance data, which again, is illustrative of a large sample size. Notwithstanding, the relatively small standard deviation for the Australia pilot study would suggest a more homogeneous group of candidates, for the small program, in comparison to both the UK and the US students.

The Netherlands will soon be administering a test with some of the same NCCPA-derived items that were administered with the UK and Australian PA programs. This third country will provide additional information for NCCPA in its evaluation of “similar” PA international programs and the viability, validity, and reliability of sharing test content derived from NCCPA. In the meantime, further validity data can be found in the following scatter plot.

Figure 1



The scatter plot shows the performance relationship between the 99 common NCCPA-derived items. The overall Pearson correlation was 0.65 indicating a moderately strong agreement. There were only 4 to 5 items that visually appear to be inconsistently measuring candidate abilities. A variety of reasons could potentially explain these results, but most likely the students were not familiar with the material and therefore should not have been asked related questions. Another possibility is

that the material was not presented in a way that was meaningful and easily understood by the students. It should also be noted, once again, that the data are limited; thus, we cannot make predictive or generalizable conclusions about the findings.

REFERENCES

1. Arbet S, Lathrop J, Hooker RS. Using practice analysis to improve the certifying examinations for PAs. *Journal of the American Academy of Physician Assistants*. 2009;22(2):31-36. [PMID 19317113]
2. De Champlain AF. Ensuring that the competent are truly competent: An overview of common methods and procedures used to set standards on high stakes examinations. *J Vet Med Educ*. 2004;31(1):61-5. [PMID 15962251]
3. Department of Health (UK). NHS: National Practitioner Program: The Competence and Curriculum Framework for the Physician Assistant (September 2006). www.dh.gov.uk