DOI: https://doi.org/10.4038/seajme.v18i2.625

Original Research Paper

Development and Validation of a Tool to assess the Retention of Clinical Anatomy Knowledge

Amaratunga, H .A[1]., Dissabandara, L[2]., Adikari, S. B[1,3]., Marambe, N. K[4].

Abstract

Introduction: A thorough knowledge in anatomy is essential for the medical student and thereafter for the practicing doctor. Therefore, retention of knowledge acquired during formative years becomes valuable. Knowledge attrition may occur as students move from preclinical years towards clinical training. A tool to measure the level of retention becomes important to assess the level of attrition and apply remedial measures.

Methodology: AKRA tool, a MCQ based tool was developed aligned to the learning objectives assessed and edited by opinions of clinicians. The AKRA tool was validated using expert opinion and statistical methods. For statistical assessment, it was administered to 100, 3rd year medical students and 20 recently passed out doctors and the scores calculated. Reliability was assessed using Cronbach alpha, and item discrimination using Pearson correlation. MCQs were graded into poor, marginal, good and excellent based on item discrimination and difficulty indices. Criterion validity was assessed by comparing the AKRA scores with those of an institutional examination. AKRA score was administered to students of two other institutions and scores were compared.

Results: AKRA tool consisted of 42 MCQs. The reliability was high (Cronbach Alph = 0.81). The average score for AKRA was 61.78 (SD=12.3) %. The AKRA scores of the students correlated well with anatomy test scores achieved at the institutional examination (r=0.76). Scores between the other institutions did not show a significant difference.

Conclusion: AKRA tool is a validated MCQ-based tool to assess the clinical anatomy knowledge in medical students and doctors.

Keywords: Anatomy, Knowledge retention, Clinical anatomy, Assessment tool

Introduction

A thorough knowledge in anatomical concepts, basically the structure of the human body seems essential for clinical examination of patients, interpretation of scans and radiological images, arriving at a proper diagnosis, performing surgeries and simple procedures such as cannulation and drawing blood (Dettmer *et al.,* 2013; Orsbon *et al.,* 2014; Vorstenbosch *et al.,* 2016). However, anatomy is classed as one of the subjects that students feel inadequately prepared for (Lawson *et al.,* 2008). Further with the reduction of time spent on dissections and teaching anatomy, clinicians and anatomists internationally have expressed concern over a perceived deficiency in anatomy knowledge of medical graduates (Drake *et al.,* 2009; Waterson & Steuart, 2005).

Many medical schools express concern about the retention of anatomy knowledge among their clinical

[1]*Department of Anatomy, Faculty of Medicine, University of Peradeniya, Peradeniya, Sri Lanka*
[2]*School of Medicine, Griffith University, Australia*
[3]*Faculty of Medicine, Manipal University College Malaysia*
[4]*Department of Medical Education, Faculty of Medicine, University of Peradeniya, Peradeniya, Sri Lanka*

Corresponding author: Dr (Mrs.) Himani A. Amaratunga
Email: himani.amaratunga@med.pdn.ac.lk

year students and have implemented various methods to enhance this knowledge. A study conducted in the United States reveals that students lack essential clinical anatomy knowledge upon entering surgical disciplines, such as general surgery and obstetrics and gynecology, consequently, they developed interactive learning modules to enhance knowledge before clerkships (Jurjus *et al.,* 2014; Jurjus *et al.,* 2016).

Dorminic *et al.,* (2017) state that anatomy knowledge attrition in their horizontally and longitudinally integrated curriculum was much less than was seen previously with a traditional curriculum (Dorminic *et al.,* 2017). Similar findings are reported by Hulme *et al.,* (2019) in musculoskeletal anatomy knowledge retention in chiropractic students (Hulme *et al.,* 2019).

Therefore, it is necessary to assess and adopt measures to improve the long-term retention of anatomy knowledge of medical students and medical professionals. Different methods have been devised in order to objectively measure the knowledge retention capabilities of students. However, there is no universally accepted method to assess knowledge retention in anatomy or any other discipline. The most commonly used method is recall and recognition in the form of labeling diagrams and multiple-choice questions developed for assessing essential anatomy knowledge (Arzi *et al.,* 1986). Many of these tools deal with assessing the retention of descriptive and recall type anatomy such as naming the carpal bones or tarsal bones, rather than clinically relevant anatomy (Lopez *et al.,* 2014; Speilmann & Oliver, 2005). While these are simple but effective tools to assess anatomy knowledge retention for students in specific fields, they may not be appropriate for every discipline.

Other researchers and educationists have developed more comprehensive tools to assess knowledge retention. One such comprehensive and in-depth anatomy competence score (ACS) has been developed by Schoeman & Chandratilake (2012) which tests subjects on theory, visual identification and clinical application and consists of MCQ and OSCE examinations. This score is comprehensive and covers all three domains equally (Schoeman & Chandratilake, 2012).

Knowledge retention needs to be assessed in context, which is the retention of useful knowledge, which can differ from country to country and will be dependent on graduate learning outcomes as described by national policies. Therefore, it is crucial to identify the knowledge deemed valuable for a specific institution or country and subsequently devise an appropriate tool to assess that knowledge. A knowledge assessment tool should be easy to administer and concise, ensuring that participants do not lose interest. Additionally, it should be designed in a manner that minimizes guessing, thereby accurately assessing the true level of knowledge. Developing such a tool is no easy task and requires in-depth knowledge and creativity. Such a tool would be extremely useful to anatomy teachers and those who create and implement new anatomy modules and curricula.

In our institution the medical course runs for 5 years and anatomy is taught formally only during the first year and no revision is done during the clinical years. Anatomy is taught on a regional basis using lectures cadaver dissections, demonstration of prosected specimens and tutorials. Clinical anatomical concepts are taught and highlighted by the anatomy teachers with occasional clinical anatomy lectures by visiting clinicians. The students sit for the institutional examination at the end of the basic science course consisting of content from anatomy, biochemistry and physiology at the end of 18 months. This examination has MCQs, Short essays, OSPEs and a viva component weighted differently. Those who achieve 50% or above in all three subjects progress to the clinical training program and continue on. The students enter clinical training in their second year with exposure to major and minor specialties in hospital wards.

With this background the objective of this study was to develop and validate a tool for the assessment of knowledge retention in applied clinical anatomy – Anatomy Knowledge Retention Assessment tool (AKRA tool).

Methods

The study was carried out at the Faculty of Medicine, University of Peradeniya from November 2017 to November 2021, with ethical approval from the Ethics

Review Committee of the Faculty of Medicine, University of Peradeniya. The tool comprises MCQs assessing applied clinical anatomy knowledge across 4 regions namely: limbs, thorax & abdomen, pelvis & perineum, neuroanatomy, head & neck. Hundred third year medical students of the Faculty of Medicine, 98 medical students from two other medical colleges and 20 doctors who had graduated within the previous year, were recruited for the validation of the tool. Participation being completely voluntary, with written informed consent. A demographic questionnaire was designed to gather information such as year of study, age and gender. The process of development is shown in figure 1.

Third year students are those who have sat for the end course examination and passed all subjects and progressed to the clinical years. The gap between the end course examination and the administration of the AKRA tool is 18 months. The cohort of doctors recruited had graduated from institutions within the country and had finished learning anatomy 4 to 5 years ago. The gap between the students and the doctors would be around 3 years. A similar group of 3rd year students were recruited from two other medical colleges in the country and there scores were compared with those of our students.

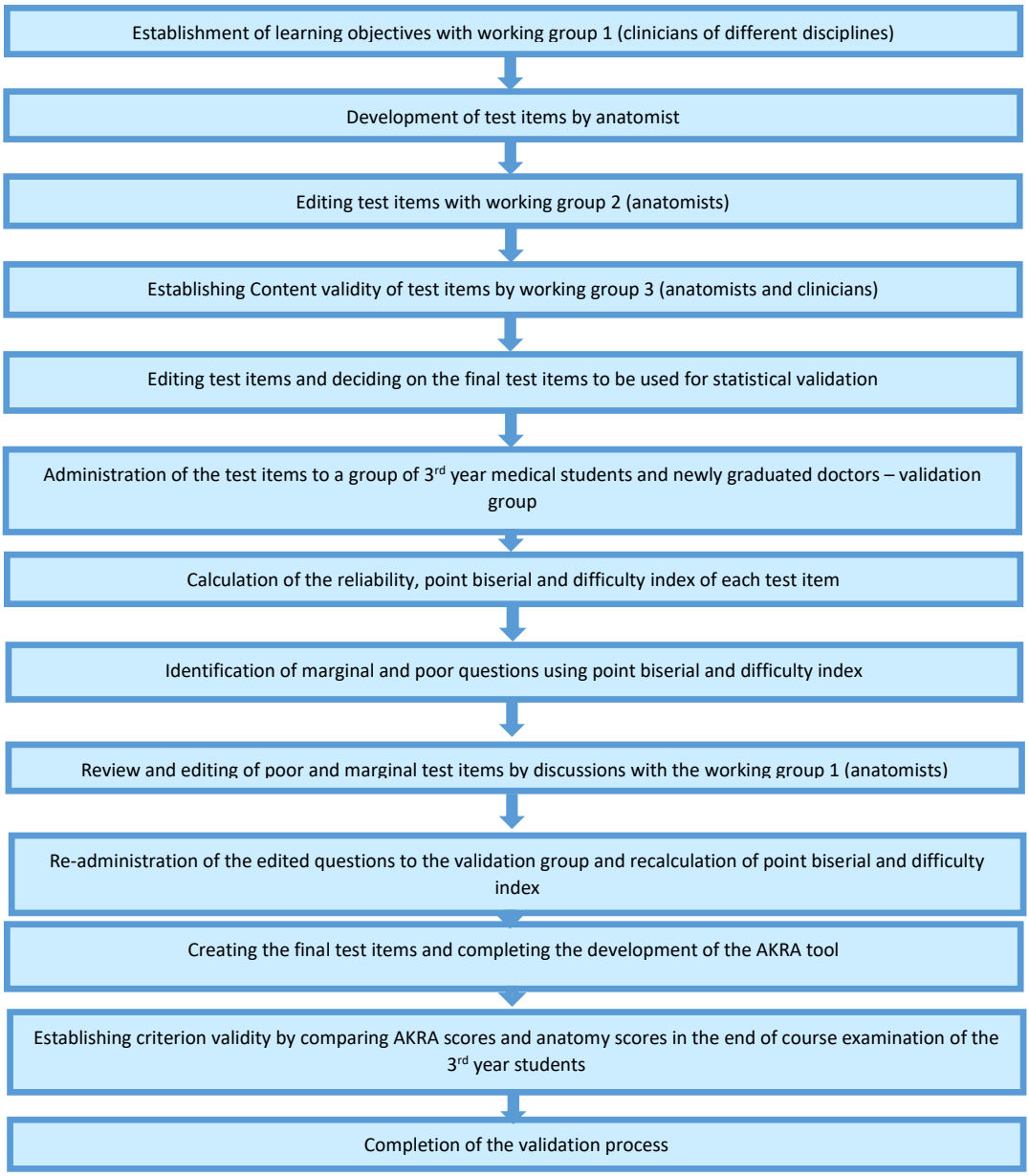

Figure 1: The process of development and validation of the AKRA tool

Establishment of learning objectives

As the anatomy curriculum follows a regional approach it consists of limbs, thorax and abdomen, pelvis and perineum and head, neck and neuroanatomy. In order to establish new learning objectives for these four anatomical regions, four working groups of clinical teachers of different disciplines were formed headed by the principal investigator who is a clinical anatomist. All 4 groups had a radiologist and a general surgeon, and in addition, an orthopedic surgeon in the working group for the anatomy of the limbs; a cardiothoracic surgeon and a gastroenterologist in the working group for anatomy of the thorax & abdomen; a gynecologist and a urological surgeon for pelvis & perineum section and a neurosurgeon, ENT surgeon and a dental surgeon in the working group for neuroanatomy, head & neck section were included. Semi- structured interviews were held with the clinical teachers in the 4 working groups where they were presented with the existing Learning Outcomes of the anatomy course at Faculty of Medicine University of Peradeniya. Following the interviews the Learning Outcomes were edited by the principal investigator based on the feedback from the working groups. The edited learning objectives were used to guide the development of the test items for the AKRA tool.

Development of the test items

Aligned to the established learning objectives, single-best-response type MCQs were developed by the principal investigator. Each question consisted of a short but adequately descriptive stem and 5 responses. The stem of some questions consisted of an image and/or a clinical vignette. Images were schematic drawings, photographs of clinical cases and images of normal and disease conditions from different imaging modalities such as ultrasound, radiography, CTs and MRIs. Clarity of the information and questions were ensured, so as to minimize misinterpretations of the presented information. The stem was followed by a question or a distinct request as to what the participant should do with the information given using their own knowledge. There was one “correct” choice, the “singe best answer” while the other four “wrong” choices were plausible distractors. The distractors were all possible answers and choices such as all of the above or none of the above were avoided. Answers were similar from the same system or region or concerning the same disease condition.

Editing test items with working group 1 (anatomists)

The developed MCQs were and discussed in a working group comprising two anatomists (working group 1). The questions were analyzed for clarity and relevance and edited accordingly. Any questions that were unclear or deemed weak were subsequently removed.

Establishing content validity

Next the MCQs were discussed in a working group comprising of two anatomists and two clinicians (working group 2) who assessed them for domain representation and relevance (Lynn, 1986). All questions found to be non-representative or non-relevant were updated and were re-reviewed by the working group.

After completion of all the questions, the tool was administered to 4 experienced anatomy teachers of the faculty who were not involved with developing the questions. Their answers were matched with those set by the principal investigator, and where any discrepancy existed, the questions were discussed and a consensus was reached.

Statistical Validation of the AKRA tool

Item analysis

The AKRA tool consisting of the chosen questions was administered to the 100 students and 20 doctors. The MCQs were marked and the AKRA scores were calculated out of 100. AKRA score were used to calculate item reliability, item discrimination, item difficulty and criterion validity.

Reliability

Internal reliability was measured using Cronbach Alph (Cronbach, 1951). It measures the internal reliability of test items; basically, how consistent the test is and how well it measures what needs to be

measured. It is calculated by taking the score from each item and correlating it with the total score for each observation. The resulting correlations are then compared with the variance for all individual item scores.

Item discrimination

For each MCQ, item discrimination was assessed using the point bi-serial correlation which is considered as a special type of Pearson correlation. It is the correlation between the right/wrong scores on a given item and the total score that the student receives when summing up their scores across the remaining items. It can also be referred to as item effect, since it is an index of an item's effectiveness at discriminating those who know the content from those who do not. Its possible range is (-1) to 1, with a minimum desired score greater than 0.15.

Item difficulty

Difficulty index was calculated by the percentage of students who got each MCQ correct. The MCQs were graded into poor, marginal, good and excellent based on item discrimination and difficulty indices as described by Ebel *et al.,* (Ebel *et al.,* 1991) (Table 1). The poor-quality questions were removed, and the marginal questions were revised. The choice of 35% as a cutoff was used to ensure that the test includes a range of item difficulties, this ensures that the test maybe used for students at different levels of knowledge. This creates a balanced assessment that can effectively measure performers of different levels.

Table 1: Grading of MCQs based on point bi-serial(Pearson correlation) & difficulty rate/index (Ebel et al., 1991)

Point bi-serial (Pearson correlation)	Difficulty rate	Grade
>0.15	>35%	excellent
>0.15	<35%	good
<0.15	>35%	marginal
<0.15	<35%	poor

Criterion validity

Criterion validity evaluates how closely the results of the test correspond to the results of a different test taken by the same cohort of students. The overall test scores for anatomy at the MBBS examination being the average score of the four, region-based end semester examinations, were taken as the established test scores, and compared with the AKRA scores of the 100, 3rd year students.

Results

Fifty test items were developed aligned to the updated learning objectives. The number of questions representing anatomy of different regions of the body were divided based on the, content and teaching hours spent. Accordingly, there were 16 questions to assess upper and lower limb anatomy, 11 questions to assess thorax and abdominal anatomy, 9 questions to assess pelvic & perineal anatomy and 14 questions to assess neuroanatomy and the head and neck region.

Content validity

During the content validation process the 50 test items were reviewed and a total of 45 questions were selected by the working group 2 (15 upper and lower limbs questions, 11 thorax and abdomen questions, 7 pelvis and perineum questions and 14 neuroanatomy, head and neck region questions).

Statistical analysis of test items

The AKRA tool consisting of 45 questions was administered to one hundred, 3rd year medical students who had finished all anatomy modules and 20 doctors who had graduated within the previous year. The average score was 61.78 (SD=12.3) %.

Reliability of the AKRA tool

The Cronbach Alph score was 0.81 indicating good reliability of the tool (Cronbach, 1951). Table 2 depicts the grading of all MCQ items based on Ebel *et al.,* (1991)(Ebel *et al.,* 1991).

Table 2*:* Difficulty index and point-biserial of test items

Anatomy of Limbs 15 items								
Question number	QL1	QL2	QL3	QL4	QL5	QL6	QL7	QL8
Difficulty index	0.52	0.23	0.47	1	0.70	0.76	0.76	0.42
Point-biserial	0.18	-0.32	0.18	-	0.75	-0.03	0.77	0.23
Quality	excellent	poor	excellent	-	excellent	marginal	excellent	excellent
	QL9	QL10	QL11	QL12	QL13	QL14	QL15	
Difficulty index	0.76	0.35	0.76	0.64	0.76	0.82	0.76	
Point-biserial	0.67	0.38	0.43	-0.21	0.22	0.17	-0.17	
Quality	excellent	good	excellent	marginal	excellent	excellent	marginal	
Thorax & abdomen 11 items								
Question number	QT1	QT2	QT3	QT4	QT5	QT6	QT7	QT8
Difficulty index	0.94	0.82	0.41	1	0.70	0.94	0.11	0.76
Point biserial	0.45	0.81	-0.44	-	-0.02	0.19	0.03	0.43
Quality	excellent	excellent	marginal	-	marginal	excellent	poor	excellent
Question number	QT 9	QT 10	QT 11					
Difficulty index	0.88	0.94	0.29					
Point biserial	0.36	0.34	0.05					
Quality	excellent	excellent	poor					
Pelvis & perineum 7 items								
Question number	QP1	QP2	QP3	QP4	QP5	QP6	QP7	
Difficulty index	0.47	0.41	1	0.88	0.58	0.64	0.64	
Point biserial	0.50	0.096	-	0.48	-0.10	0.31	-0.016	
Quality	excellent	marginal	-	excellent	marginal	excellent	marginal	
Neuroanatomy head & neck 12 items								
Question number	QN1	QN2	QN3	QN4	QN5	QN6	QN7	QN8
Difficulty index	0.29	0.76	0.76	0.76	0.76	0.82	0.35	0.82
Point biserial	0.23	0.52	0.28	0.021	0.87	0.54	0.097	0.34
Quality	good	excellent	excellent	marginal	excellent	excellent	marginal	excellent

	QN9	QN10	QN11	QN12
Difficulty index	0.58	0.70	0.82	1
Point biserial	0.34	0.31	0.17	-
Quality	excellent	excellent	excellent	-

Four questions were answered correctly by all participants, and a point bi-serial could not be calculated. Therefore, these questions were not graded. However, considering that the purpose of the AKRA tool is to assess anatomy knowledge retention rather than to discriminate between good and bad students, and because the panel identified the covered area in these MCQs as essential knowledge, it was decided to include these questions in the final tool.

In the limbs section, one poor quality question and 4 marginal questions were identified while in the thorax & abdomen section, 1 poor quality question and 3 marginal questions were identified. In the pelvis & perineum section, 3 marginal questions were identified and, in the neuroanatomy, head & neck section, 2 marginal questions were identified.

Re-evaluation of poor and marginal questions

Out of the 45 questions evaluated 3 questions that fell into the poor category were removed, while 10 questions which fell into the marginal category were re-evaluated by working group 2 consisting of anatomists and clinicians and the questions were modified or removed. The modified questions were administered to the same group and the difficulty index and the point bi-serial (Pearson correlation) were recalculated. The difficulty index and point biserial of the revised questions are given in table 3.

Table 3: Difficulty index and point bi-serial (Pearson correlation) after modifications

	Previous values	Recalculated value	Previous values	Recalculated value	Previous values	Recalculated value
Limbs	QL6	QL6	QL12	QL12	QL15	QL15
Difficulty index	0.76	0.65	0.64	0.56	0.76	0.6
Point-biserial	-0.03	0.14	-0.21	0.13	-0.17	0.23
Quality	marginal	good	marginal	good	marginal	excellent
Thorax & abdomen	QT3	QT3	QT5	QT5		
Difficulty index	0.41	0.55	0.70	0.75		
Point-biserial	-0.44	0.16	-0.02	0.18		
Quality	marginal	excellent	marginal	excellent		
Pelvis & perineum	QP2	QP2	QP5	QP5	QP7	QP7
Difficulty index	0.41	0.42	0.58	0.34	0.64	0.70
Point-biserial	0.096	0.17	-0.10	0.15	-0.016	0.16
Quality	marginal	excellent	marginal	good	marginal	excellent

Neuroanatomy head & neck	QN4	QN4	QN7	QN7
Difficulty index	0.76	0.78	0.35	0.58
Point-biserial	0.021	0.16	0.097	0.15
Quality	marginal	excellent	marginal	excellent

Criterion validity

To check criterion validity the AKRA score of the 42 questions, at the 1st attempt in the 100 students were compared with the MCQ scores each one of them achieved for the end of course institutional anatomy examination (mean score 64.1(SD=9.7). A significant positive correlation (r=0.76, p<0.0001) was observed between the AKRA scores and the MCQ scores of the institutional anatomy examination (Figure 2).

The final AKRA tool consisted of 42 MCQs. (Limbs - 14, Thorax & Abdomen – 9, Pelvis & Perineum- 7 and Neuroanatomy and Head & neck-12). Each question was tagged with a question number and the level of Blooms taxonomy which was decided with the consensus of working group 2 (Blooms *et al*., 1956). There were 10 questions of Blooms level 1- 'remember', 19 questions of level 2- 'understanding' and 13 questions of level 3 - 'application'.

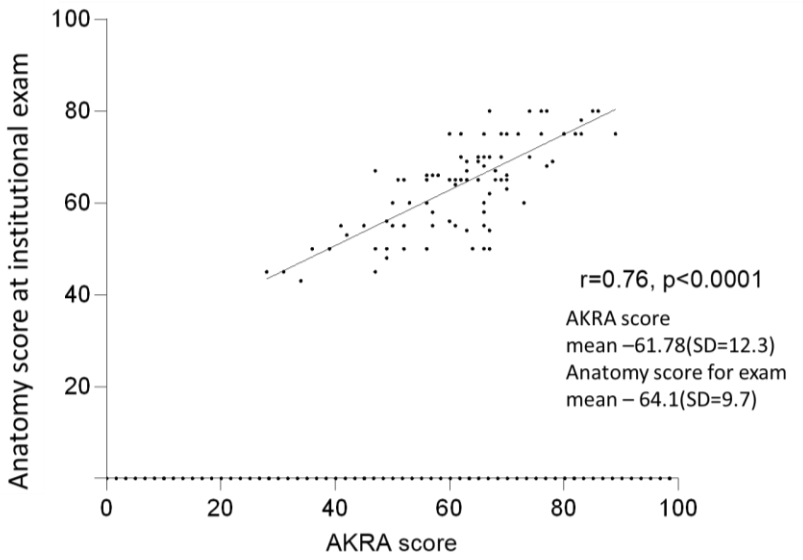

Figure 2: Correlation between the AKRA score and MCQ scores at the institutional anatomy examination

AKRA scores of medical students from other Universities

The AKRA tool were administered to 3rd year students of two other state Universities. It was sent to 150 students from each institute. Fifty-two from one institute and 46 from the other (98 students) responded (response rate – 65.3%). The AKRA score obtained by them were 61.19% (SD=11.56) and 60.7% (SD=11.2). These values were compared to the scores obtained by 3rd year students from our institute using One-way ANOVA. There was no significant difference in the scores between the three institutions [F = 1.096 (2, 174), p = 0.33].

Discussion

A strong understanding and retention of clinical anatomy knowledge are crucial for practicing doctors. Forgetting the essential elements of anatomy, can

have a negative impact on their clinical practice (Jurjus *et al.,* 2014, Jurjus *et al.,* 2016). To address this issue, the development of a tool to measure the retention of clinical anatomy knowledge becomes imperative. We have created a short MCQ-based tool to assess clinical anatomy, covering all regions of the body, and have validated this tool.

The concept of developing a universal tool to evaluate the retention of anatomy knowledge has been present for a considerable period. Nonetheless, there is no widely accepted tool established as the gold standard for this purpose. Existing literature highlights two straightforward tests employed to assess the retention of musculoskeletal anatomy knowledge: identifying carpal bones and tarsal bones on a provided diagram. These tests have been utilized for over two decades to evaluate knowledge retention in both medical and paramedical fields. However, the narrow coverage of content and the assessment of basic gross anatomy knowledge limited to the spheres of recalling, remembering, and identifying deem these tools less than ideal for evaluating knowledge retention in medical students. The AKRA tool primarily emphasizes clinical and practice-oriented anatomy, as this is essential for clinical year students and practicing doctors to retain. It integrates concepts of gross anatomy and radiological anatomy into the multiple-choice questions (MCQs). Additionally, the AKRA tool includes questions from all anatomical regions, thereby evaluating knowledge related to the entire body. While assessing recall abilities, it also tests the application of knowledge.

The AKRA tool, is a short and easily administered instrument with the added advantage of compatibility with most available survey software. It can be seamlessly integrated into other software, administered via email, or shared on social media, rendering it highly versatile. An additional benefit is to, and representative of the targeted construct it is designed to measure (Rusticus & Lovato, 2014). It plays a vital role in the development of any new test tool and provides evidence about the validity of such by assessing the degree to which the test tool measures the targeted content (Anastasia, 1988). This enables the results obtained by administering such a tool to be used to make meaningful and appropriate inferences and decisions (Messick, 1989, Moss, 1995). All three components of content validity: the ability to incorporation of answers to the multiple-choice questions (MCQs) if needed. This will allow participants to understand why their answers were incorrect and grasp the underlying concepts. Moreover, when administered through Google Forms or Microsoft Forms, the AKRA score is automatically calculated and presented at the end after the form has been submitted, enabling participants to assess their own level of knowledge.

The decision to include MCQ type test items in the AKRA tool was based on the fact that MCQ questions even though probing, takes less time to answer, is more objective, eliminating subjective effects of marking, gives total parity across students and can be marked electronically and the scores given at the end of the test easily (Bandaranayake, 2018). Among the factors against using MCQs as an assessment tool, guessing stands out as a significant challenge. Mitigating this issue can be challenging, but one solution is to provide five alternatives, aiming to minimize the impact of guessing. Negative marks are typically not assigned in single-response questions, as each question evaluates a distinct knowledge base. While these aspects may diminish the effectiveness of the MCQ as an assessment tool, the AKRA tool serves a slightly different purpose. Designed to assess the retention of anatomy knowledge rather than the knowledge base or the effectiveness of learning, it can be considered a suitable method in this context (Roediger & Marsh, 2005). Single response questions were used since this format is familiar to students and doctors alike.

In the process of validating the developed tool, expert opinion as well as statistical methods were utilized. Content validity was assured by employing the well-established and reliable anatomy content of the first- and second-year curriculum at the University of Peradeniya, Sri Lanka. Content validity refers to the degree to which an assessment instrument is relevant domain definition, domain representation, and domain relevance were tested (Sireci, 1998). Content domain was defined as the subject matter taught during the preclinical years at Peradeniya. Learning objectives created and modified over the years by numerous subject experts were used as the content domain. Since the objectives were based on universally accepted anatomy textbooks and concepts, with the subject experts being anatomy teachers with many years of experience in the field,

the defined content domain can be considered as valid. Further the objectives currently in use were discussed with clinical teachers and their ideas and comments were incorporated into them. This enriched the existing learning objectives and gave more credibility to an already established system, strengthening its validity further.

Since the earliest years of educational testing, validation procedures have sought to demonstrate the criterion validity of a test by correlating test scores with an external criterion (Bingham, 1937; Kelly, 1927; Thurstone, 1932). In this context, the time-tested MBBS examination serves as the external criterion for our study. Conducted by the Department of Anatomy, this examination's questions MCQs, essays and OSPEs are formulated and scrutinized by anatomy teachers, making the scores achieved a valid external criterion for assessing criterion validity. A strong correlation between the AKRA score and anatomy scores at the MBBS implies good criterion validity, which can be interpreted as a correlation between academic performance and knowledge retention as well. High achievers are more likely to retain knowledge acquired for a longer duration.

The AKRA tool encompassed questions spanning three Bloom's levels: remembering, understanding, and application. The largest proportion of questions (45%) pertained to the understanding level, with 31% focusing on the application level. The majority of questions were centered around clinical cases and involved important anatomical concepts with clinical significance. This emphasis on clinically relevant basic science knowledge is crucial for both students and doctors to ensure safe medical practice. Notably, students exhibited similar performance across questions of various Bloom's levels. Questions assessing knowledge application and understanding were answered as effectively as those testing mere recall. This is an encouraging observation, countering concerns raised by some researchers. They have suggested that poor performance in anatomy, despite question difficulty comparable to other disciplines, might stem from a tendency among students to prioritize memorization over understanding during the learning process (Miller, 1962).

The AKRA tool covering all anatomical regions is a short MCQ-based tool that helps assess the retention of anatomy knowledge in both medical students and doctors. The comparability of the AKRA score between institutions ascertains its generalizability. However, the AKRA tool is not recommend to other health care professionals since the depth of knowledge tested in this tool is specific for the medical profession. It is hoped that it will aid anatomists to assess new teaching methods and curricular changes by testing anatomy knowledge retention in different situations and comparing them.

One of the limitations of the AKRA is that only MCQs were used, allowing a chance for guessing and difficulty in reusing the tool especially within short retention intervals. Further, only 42 questions were used and even though all regions were covered it may not be a comprehensive coverage. The clinically oriented questions are of a higher difficulty level, and therefore the tool can only be used for assessment of more mature medical students and doctors, and it cannot be adopted for other health science students or practitioners. However, the tool maybe customized to suite different levels of students and professionals by modifying the questions and adjusting the difficulty level. Therefore, several different variations of the tool can be made available.
can serve as model for other medical schools to modify according to their institutional needs.

Conclusion

The MCQ-based AKRA tool designed to evaluate the retention of anatomy knowledge, has undergone rigorous development and validation processes, incorporating expert opinions and statistical methodologies. Its robust validation and concise nature support its recommendation for efficient assessments of retention of clinical anatomy among both medical students and doctors. Further this tool

Declaration: Ethics approval was granted by the institutional review board. Written informed consent was sort from all participants.

Availability of data and materials: The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

The authors declare no competing interests relevant to the content of this article.

No funding was received for conducting this study.

Authors' contributions: HAA was involved in study design, implementation, data collection, data analysis and interpretation, major contribution to writing of manuscript, while DLOD was also involved in study design, implementation, data analysis and interpretation, major contribution to editing of manuscript, SBA and KNM too was involved in study design, implementation, data analysis and interpretation, editing of manuscript. All authors approved the final manuscript.

References

Anastasia, A. (1988). Psychological testing. New York: Macmillan Publishing.

Arzi, H. J., Ben-Zvi, R., Ganiel, U. (1986). Forgetting versus savings: The many facets of long-term retention. Anatomical Science Education, 70(2), 171–188.

Bandaranayake, R.C. (2008). Setting and maintaining standards in multiple choice examinations: AMEE Guide No. 37. Medical Teacher, 30(9-10), 836-845.

Bingham, W.V. (1937). Aptitudes and Aptitude Testing. New York: Harper.

Bloom, B.S., Englehart, M.D., Furst, E.J., Hill, W.H., Krathwohl, D.R. (1956). Taxonomy of Educational Objectives: Handbook I, Cognitive Domain. New York: David McKey.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. Psychomerika, 16, 297-334.

Dettmer, S., Schmiedl, A., Meyer, S., Giesemann, A., Pabst, R., Weidemann, J., Wacker, F.K., Kirchhoff, T. (2013). Radiological anatomy - evaluation of integrative education in radiology. RöFo, 185(9), 838-843.

Doomernik, D.E., van Goor H, Kooloos JGM, Ten Broek RP. (2017). Longitudinal retention of anatomical knowledge in second-year medical students. Anatomical Science Education, 10(3), 242-248.

Drake, R.L., McBride, J.M., Lachman, N., Pawlina, W. (2009). Medical education in the anatomical sciences: the winds of change continue to blow. Anatomical Science Education, 2(6), 253–259.

Ebel, R..L, Frisbie, D.A. (1991).Essentials of Educational Measurement. Englewood Cliffs, NJ: Prentice Hall.

Hulme, A.K., Luo, K., Štrkalj, G. (2020). Musculoskeletal Anatomy Knowledge Retention in the Macquarie University Chiropractic Program: A Cross-Sectional Study. Anatomical Science Education, 13(2), 182-191.

Jurjus, R.J., Lee, Ahle, S., Brown, K.M., Butera, G., Goldman, E.F. Krapf, J. (2014). Anatomical Knowledge Retention in Third-Year Medical Students Prior to Obstetrics and Gynecology and Surgery Rotations. Anatomical Science Education,**7**(6), 461-468.

Jurjus, R.A., Brown, K., Goldman, E., Galoosian, A., Butera, G., Krapf, J.M. (2016). Curricular Response to Increase Recall and Transfer of Anatomical Knowledge into the Obstetrics/Gynecology Clerkship. Anatomical Science Education, 9, 337–343.

Kelley, T. L. (1927). Interpretation of Educational Measurement. New York: World Book Company.

Lawson, M., Jolly, B., Benjamin, J. (2008). The Australian medical education study. What makes for success in medical education? Government report. Department of Education, science and training. *ACT*: Commonwealth of Australia: Canberra.

Lopez, J.M.C., Diaz-Mancha, J.A., Heredia-Rizo, A.M., Fernandez-Seguin, L.M., Padilla, J.P., Dominguez-Maldonado, G., Munuera, P.V. (2014). The Tarsal Bone Test: A Basic Test of Health Sciences Students' Knowledge of Lower Limb Anatomy. Biomed research International, 1-6.

Lynn, M.R. (1986). Determination and quantification of content validity. Nursing Research, 35(6), 382-385.

Messick, S. (1989). Validity in Educational measurement. New York: American Council on Education.

Miller, G.E. (1962). An inquiry into medical teaching. Journal of Medical Education, 37(3), 185–191.

Moss, P.A. (1995). Themes and variations in validity theory. Educational Measurement: Issues and Practice, 14, 5–13.

Orsbon, C.P., Kaiser, R.S., Ross, C.F. (2014). Physician opinions about an anatomy core curriculum: a case for medical imaging and vertical integration. Anatomical Science Education, 7(4), 251-261.

Roediger, H.L., Marsh, E.J. (2005). The positive and negative consequences of multiple-choice testing. Journal of Experimental Psychology. Learning, Memory, and Cognition, 31, 1155-1159.

Rusticus, S.A., Lovato, C.Y. (2014). Impact of Sample Size and Variability on the Power and Type I Error Rates of Equivalence Tests: A Simulation Study. Practical Assessment, Research & Evaluation, 19, 1-10.

Schoeman, S., Chandratilake, M. (2012). The anatomy competence score: a new marker for anatomical ability. Anatomical Science Education, 5(1), 33-40.

Sireci, S.G. (1998). The construct of content validity. Social Indicators Research, 45, 83–117.

Spielmann, P.M., Oliver, C.W. (2005). The carpal bones: a basic test of medical students and junior doctors knowledge of anatomy. Surgery, 3(4), 257–259.

Thurstone, L.L. (1932). The Reliability and Validity of Tests. Michigan: Edwards Brothers.

Vorstenbosch, M.A., Kooloos, J.G., Bolhuis, S.M., Laan, R.F. (2016). An investigation of anatomical competence in junior medical doctors. Anatomical Science Education, 9(1), 8-17.

Waterson, S.W. Stewart, I.J. (2005). Survey of clinician's attitudes to the anatomical teaching and knowledge of medical students. Clinical Anatomy,18, 380–384.