Hips Don't Lie: Expert Opinions Guide the Validation of a Virtual 3D Pelvis Model for Use in Anatomy Education and Medical Training.

Edgar R. Meyer*, Amber M. James and Dongmei Cui
*Corresponding Author: emeyer@umc.edu

HAPS Educator. Vol 22 (2), pp. 105-118. Published August 2018.
doi: 10.21692/haps.2018.023

Meyer E.R. et. al. (2018). Hips Don't Lie: Expert Opinions Guide the Validation of a Virtual 3D Pelvis Model for Use in Anatomy Education and Medical Training. *HAPS Educator* 22 (2): 105-118. doi: 10.21692/haps.2018.023

Hips Don't Lie: Expert Opinions Guide the Validation of a Virtual 3D Pelvis Model for Use in Anatomy Education and Medical Training

Edgar R. Meyer, M.A.T.[1,3]; Amber M. James, M.C.A.[2]; and Dongmei Cui, M.D. (hon), Ph.D.[1,2,3]
[1]School of Graduate Studies in the Health Sciences, University of Mississippi Medical Center
[2]School of Medicine, University of Mississippi Medical Center
[3]Department of Neurobiology & Anatomical Sciences, Clinical Anatomy Division, University of Mississippi Medical Center, 2500 N. State St., Jackson, MS 39216
Corresponding Author: Edgar R. Meyer, emeyer@umc.edu

Abstract
Virtual three-dimensional (3D) anatomical models have become popular in the education of students in graduate and undergraduate-level anatomy courses, including anatomy and physiology. There is a need for more research on the effectiveness of these models on student learning, especially in anatomy and physiology, and on the validation of the models implemented in anatomy education. This study focuses on the development of a list of criteria for validating 3D models of the pelvis, using a four-round Delphi method. The Delphi method successfully confirmed the validity of the final list of pelvic criteria, each item with an item content validity index (I-CVI) ≥ 0.83. The average scale content validity index (S-CVI/Ave) of the entire list was calculated to be rigorously valid (0.92). These valid pelvic criteria will be used in future studies to inform additions to an existing virtual 3D model of the pelvis and validate completed virtual 3D pelvic models. https://doi.org/10.21692/haps.2018.023

Key words: anatomy education, virtual 3D models, pelvic models, Delphi method, medical training

Introduction
Anatomy education has undergone many changes throughout its history. One such change has been the implementation of virtual 3D anatomical models in teaching diverse cohorts of anatomy students. Virtual 3D anatomical models have been shown to be beneficial to students of anatomy in both medical (Qayumi *et al.* 2004, Nicholson *et al.* 2006, Brown *et al.* 2012, Cui *et al.* 2017) and dental programs (Maggio *et al.* 2012). Thus, virtual 3D anatomy models of structures in intricate or small areas of the body can be used to improve medical student retention of the corresponding anatomical information.

In addition, virtual 3D anatomical models show promise in undergraduate anatomy education in both undergraduate medicine (Hisley *et al.* 2008, Marsh *et al.* 2008, Müller-Stich *et al.* 2013) and kinesiology (Hoyek *et al.* 2014). Womble (1999) showed that undergraduate students in an anatomy and physiology course favored the use of virtual 3D anatomy, but more research is needed to determine the impact of virtual 3D anatomy on undergraduate education in anatomy and physiology.

In medical education, 3D anatomical models are frequently used to help medical trainees master human anatomy. In the clinical setting, virtual 3D models are used to guide residents and physicians in devising optimal surgical approaches, accessing specific anatomical regions, and excising or repairing structural pathology. These virtual 3D models are especially useful for elucidating complex anatomical regions, such as the pelvis which is a confined space with limited access via dissection. A few studies describe the construction of virtual 3D pelvic models (Beyersdorff *et al.* 2001, Parikh *et al.* 2014, Sergovich *et al.* 2010), and one in particular describes the impact of physical and virtual 3D pelvic models on medical education (Khot *et al.* 2013). However, the validation of pelvic models has not been reported. A review study by Azer and Azer (2016) asserts a need for reporting the validation of 3D models and the assessment instruments used in educational studies for evaluating their impact on learning.

Since there are variations among 3D models, there is a need for consistent, valid anatomical models. In order to ensure the creation of valid anatomical models, an assessment instrument for measuring the validity of 3D models such as the pelvis can be generated using expert opinions. One of the first steps in acquiring expert opinions utilizes the Delphi method, which involves administering a series of two or more rounds of questionnaires to a team of experts in a particular field or discipline (Dalkey and Helmer 1963, Helmer 1967, Caves 1988, Hasson and Keeney 2011, Lisk *et al.* 2014). Lisk *et al.* (2014) used the Delphi method to establish a list of criteria for anatomical structures to include in a curriculum of musculoskeletal anatomy for training residents in Physical Medicine and Rehabilitation.

The validity and reliability of the instrument itself needs to be considered in the process of gathering information for its construction. Hasson and Keeney (2011) cite a number of studies emphasizing the fact that there is controversy over the validity and reliability of the Delphi method. There are also sources that assert the validity and reliability of the Delphi method (Helmer 1967, Caves 1988).

continued on next page

Although Hasson and Keeney (2011) mention a number of different types of validity and reliability to consider in the use of the Delphi method, one in particular known as content validity is highly emphasized in this study. A number of articles have demonstrated the importance of considering the content validity in the construction of scales or instruments (Lawshe 1975, Waltz and Bausell 1981, Lynn 1986, Davis 1992, Polit and Beck 2006). Polit and Beck (2006) offer a number of suggestions for the reporting of content validity, including the importance of differentiating between item-level and scale-level content validity. While the former refers to the validity of the information within each item on an instrument, the latter refers to the overall validity of the entire instrument. Determining content validity on both levels requires the calculation of a value known as the content validity index (Lawshe 1975, Polit and Beck 2006), and the calculations for both the item-level content validity index (I-CVI) and the scale-level content validity index (S-CVI) are described by Polit and Beck (2006). The item-level and scale-level content validity will be the primary focus of this paper as they are expressly relevant to the construction of criteria lists of pelvic anatomical structures that are important to consider in teaching pre-health and medical students, as well as medical residents.

The purpose of this article is to describe in detail the Delphi method used to develop a list of valid criteria for validating 3D anatomical pelvic models, the steps that will be taken to generate additional structures to incorporate into an already existing virtual 3D pelvic model based on experts' feedback, and the target audiences for such a model. In addition, this article aims to address the lack of reporting of the validity of created anatomical models and assessment instruments in the literature.

Methods

Context

This study was conducted at the University of Mississippi Medical Center (UMMC), a large, urban academic medical center in the southeastern United States. This institution serves as the only academic medical center in the state, making it a critical entity for educating most of the state's healthcare professional students in a number of programs, including medicine, dentistry, nursing, pharmacy, graduate studies, population health, and other allied health fields. Several of these disciplines rely heavily on anatomical education that consists of traditional lectures, laboratory experiences, and active learning sessions. Virtual 3D anatomy has been implemented specifically in the medical gross anatomy and medical neuroscience and behavior courses as independent, voluntary 3D learning sessions, but these learning experiences have not yet been fully integrated into the anatomy courses. This study will describe steps toward validating a set of virtual 3D anatomical models through the development of valid criteria for a pelvic model. The Delphi method was implemented in four rounds to gather expert opinions for developing the list of criteria. The particular Delphi procedure used was a classical method in the sense that it incorporated more than three rounds of administrations beginning with an open-ended, qualitative session (Hasson and Keeney 2011). However, similar to the modified Delphi design, the expert responses were collected in the subsequent rounds using a variety of methods (Hasson and Keeney 2011), including in-person interviews, online communications, and paper-based surveys.

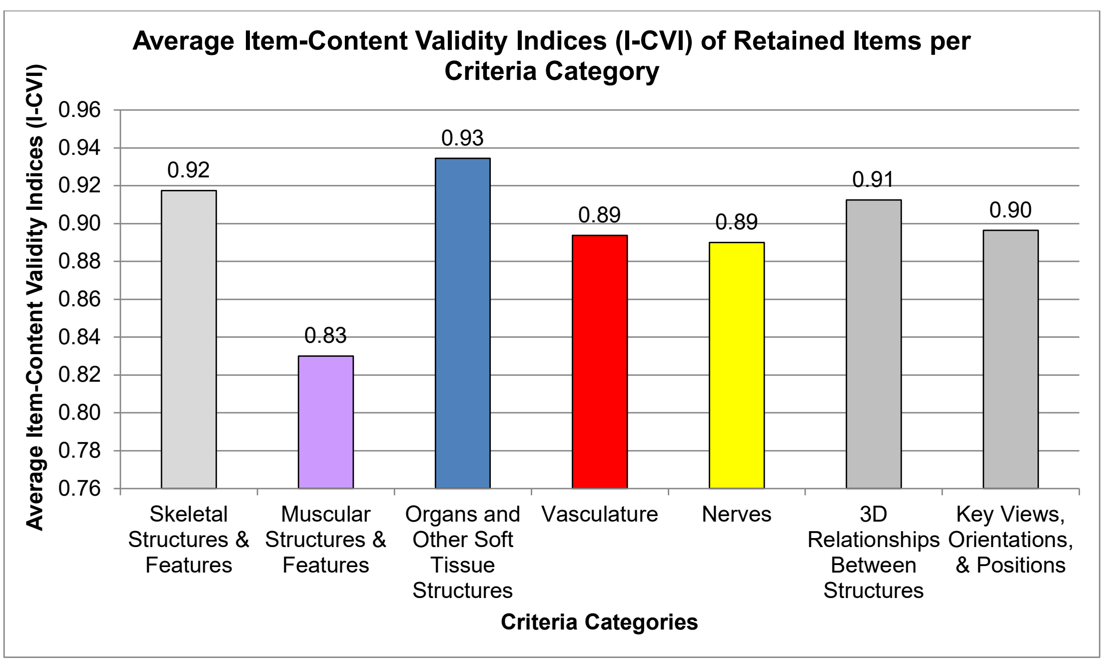

Figure 1. Average Item-Content Validity Index (I-CVI) Values for the Retained Items in the Final Criteria List. These average I-CVI values were calculated from the average all I-CVI values for items within each category in the final criteria list.

continued on next page

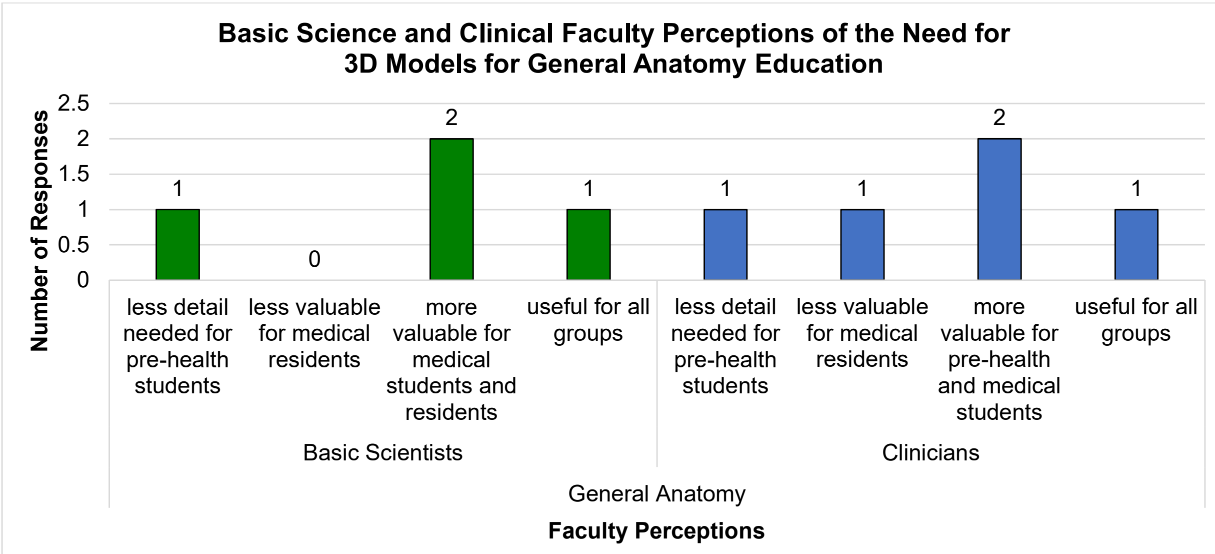

Figure 2. Faculty Perceptions of 3D Models in General Anatomy Education. Even though there are a total of 7 experts in this study, this graph shows a total of 9 responses (4 from basic science faculty and 5 from clinical faculty) because one scientist and one clinician provided two responses to the relevant survey question regarding their perception of the need for 3D models in teaching general anatomy to pre-health students, medical students, and medical residents. More basic scientists perceived general 3D anatomy to be more valuable for medical students and residents than pre-health students while more clinicians perceived general 3D anatomy to be more valuable for pre-health and medical students than medical residents.

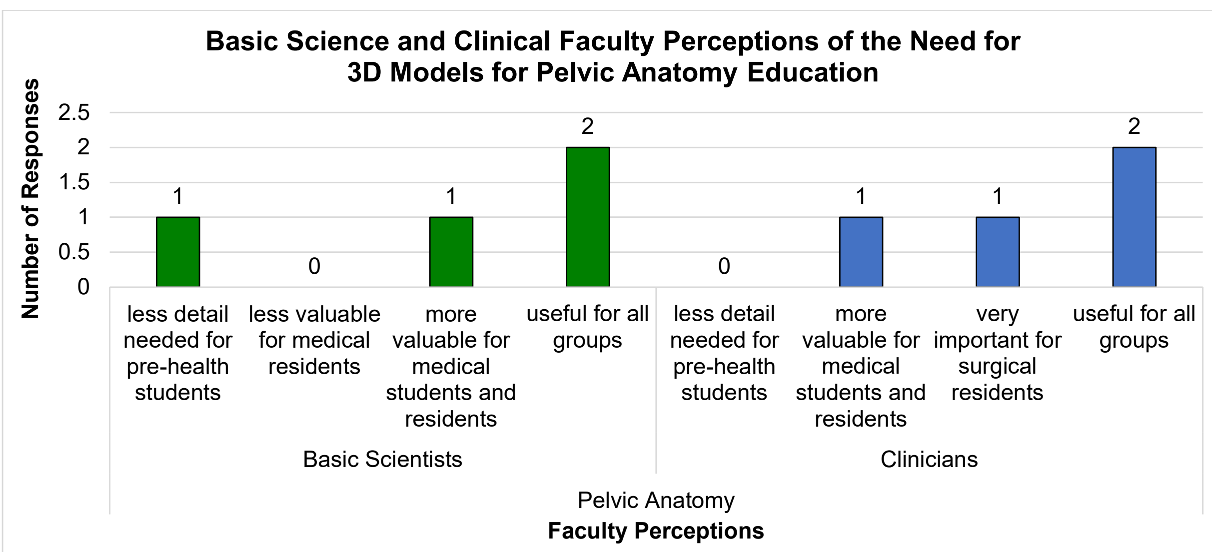

Figure 3. Faculty Perceptions of 3D Models in Pelvic Anatomy Education. Even though there are a total of 7 experts in this study, this graph shows a total of 8 responses (4 from basic science faculty and 4 from clinical faculty) because one clinician provided two responses to the relevant survey question regarding their perception of the need for 3D models in teaching pelvic anatomy to pre-health students, medical students, and medical residents. More basic scientists and more clinicians perceived pelvic 3D anatomy to be useful for all three learning groups.

continued on next page

Table 1. The Number of Experts Solicited and Recruited for the Delphi Method.

Selection of Experts				
Expert Category	**Expert Discipline**	**# Solicited (n)**	**# Recruited (n)**	**# Dropped (n)**
Basic Scientists	Anatomy	5	3	0
Clinicians	Obstetrics and gynecology	11	1	0
	Pediatric urology	2	1	
	Urogynecology	1	1	0
	Urology	4	1	1 (after round 2)
	Total	**23**	**7**	**1**

This table displays the number of clinical and basic science experts who were solicited to participate in the study and the number of experts who submitted responses. Experts were solicited and recruited from faculty members in the anatomical sciences and from clinicians in the specialties or subspecialties of obstetrics and gynecology, pediatric urology, urogynecology, and urology. The urologist dropped out of the study, leaving a total of six experts to complete the study in full.

The following tables are available on the HAPS website HERE.

1. *Table 2: First generated list of expert responses (Delphi method rounds 1 and 2).*
2. *Table 3: Expert ratings and content validity indices (Delphi method round 3).*
3. *Table 4: Expert ratings and content validity indices (Delphi method round 4).*
4. *Table 5: A list of items that were not in agreement between rounds three and four.*
5. *Table 6: A list of items that were removed from the final criteria list.*
6. *Table 7: The final list of retained items and their content validity.*

continued on next page

Table 8. Overall Internal Consistency Within Rounds 3 and 4.

Round	Total Number of Items (n)	Number of Items Excluded	Number of Valid Items	Overall Cronbach's Alpha Values
Round 3	159	23	136	0.970
Round 4	159	29	130	0.984

Coefficient Alpha values have the following reliability considerations:
>0.9 = excellent, >0.8 = good, >0.7 = acceptable, >0.6 = questionable, >0.5 = poor
(George and Mallery 2003).
The items which were excluded from the total number of items within each round were removed because one or more of the experts within each round did not provide a rating for the items on the four-point Likert-type scale.

Table 9. Item-Specific Internal Consistency Within Rounds 3 and 4.

Round	Item-Specific Cronbach's Alpha Values	Valid Item Numbers (based on Table 2)	Number of Valid Items (n)
Round 3	0.968	75, 76, 86, 100	4
	0.969	5, 17, 19-21, 30-33, 36-40, 42, 46, 47, 55, 58, 62, 65-70, 72, 77, 84, 95-97, 99, 110-117, 120-124, 129-133, 135, 136, 138, 139, 150, 152, 169-173	62
	0.970	7-9, 14, 24, 25, 28, 44, 49, 56, 64, 82, 87, 89-91, 98, 134, 137, 143-145, 149, 162, 165, 164, 168	27
	0.971	23, 27, 29, 41, 88	5
		Total →	98
Round 4	0.983	40, 46, 47, 49, 50-52, 55, 61, 65, 69-71, 74, 76, 80, 86, 99-104, 106, 113-115, 135-138, 140, 167	33
	0.984	5, 7-12, 17-21, 24-28, 30-33, 36-39, 41-45, 48, 59, 60, 64, 66-68, 73, 75, 77, 82-84, 87-91, 98, 105, 107, 110-112, 116, 117, 120-124, 130-134, 139, 143, 149, 152, 162, 166, 168, 169	74
	0.985	22, 23	2
		Total →	109

Coefficient Alpha values have the following reliability considerations:
>0.9 = excellent, >0.8 = good, >0.7 = acceptable, >0.6 = questionable, >0.5 = poor
(George and Mallery 2003).
The item numbers not present in the table represented those items which were excluded from the total number of items within each round because those items had zero variance, because one or more of the experts within each round did not provide a rating for the items on the four-point Likert-type scale, or because the item numbers represented the additional spaces provided in Table 2 for experts to write in items deemed missing from the scale.

continued on next page

Table 10. Internal Consistency Between Experts' Ratings for Rounds 3 and 4.

Expert	Total Number of Items (n)	Number of Items Excluded	Number of Valid Items	Overall Cronbach's Alpha Values
Expert 1	159	22	137	0.917
Expert 2	159	4	155	**0.666**
Expert 3	159	18	141	0.832
Expert 4	159	3	156	0.740
Expert 5	159	4	155	**0.672**
Expert 6	159	26	133	0.856

Coefficient Alpha values have the following reliability considerations:
>0.9 = excellent, >0.8 = good, >0.7 = acceptable, >0.6 = questionable, >0.5 = poor
(George and Mallery 2003).
Coefficient Alpha values in red are considerably inconsistent with the other values.
Experts 1 through 3 include clinicians while experts 4 through 6 include basic scientists.

Participants

The participants in this study were experts, including faculty (physicians and scientists) who have proficiency pertaining to pelvic anatomy and who are involved in educating medical trainees on this anatomical region. These experts were selected, following specific guidelines, including the development of a list of disciplines pertinent to the study, the identification of personnel in those disciplines, the solicitation of those personnel to participate in the study, the classification of the personnel based on their qualifications, and the recruitment of the expert personnel in classification order (Okoli and Pawlowski 2004). The experts remained anonymous to one another throughout the entirety of the study. A total of 23 experts were recruited, and from this group a total of seven experts were elected to participate in the study. One expert dropped out of the study after the second round, leaving six experts to complete the study in full. While Okoli and Pawlowski (2004) suggested an expert pool of ten to eighteen individuals, Lynn (1986) suggested that an expert pool greater than ten was not necessary. The greater the number of experts, the greater the difficulty in establishing consensus among them. Table 1 provides details on the experts' respective disciplines. The protocol for this study was approved by the Institutional Review Board (IRB) of the University of Mississippi Medical Center (IRB # 2017-0220), and informed consent was obtained from all participants.

The Delphi Method

In the first round, experts in the Delphi panel were administered a list of free-response survey items. These items included questions inquiring about the experts' opinions on the use of 3D technology in teaching anatomy in general, and in teaching anatomy to pre-health students, medical students, and medical residents. Expert opinion was also sought in teaching the anatomy of the pelvis to the same three cohorts. Additionally, experts were asked for their opinion on what anatomical structures and three-dimensional relationships are most important to include in creating 3D models of the pelvis for the purpose of educating pre-health students, medical students, and medical residents. A few examples of these latter questions include the following items:

1. What anatomical structures, in your professional opinion, are most important to be included in a 3D anatomical model of the pelvis?
2. What 3D relationships among anatomical structures are most critical to portray to students and trainees through a 3D anatomical model of the pelvis?
3. Are there any key views, orientations, and/or positions of the anatomical structures within the pelvis that are vital for medical students to learn? If so, please describe them.

Several weeks before the questionnaires were distributed to the experts, three internal faculty members in the Clinical Anatomy Division reviewed the questions and provided their feedback, vetting the survey in its readiness for dissemination. One of the faculty reviewers later also served as an expert on the Delphi panel, but this individual was not aware of potential recruitment and selection as an expert for this study during the questionnaire review process. After the questionnaires were distributed to the experts, the experts' opinions were solicited using both email and in-person formats. The interviews were recorded manually on paper copies of the questionnaire and transcribed electronically.

continued on next page

In the second round, the transcriptions from the personal and phone interviews were submitted to each corresponding interviewee via email communications to confirm their responses from the first round. However, expert respondents who chose to submit their responses to the survey in round one via email were able to forego the second round since there was no need to confirm the validity of their direct written replies. Each set of revised transcriptions and first-round email submissions were then analyzed. All nonrepeating anatomical structures, 3D relationships, and key views, orientations, and positions were included in a generated list of tentative pelvic criteria. This generated list is provided in Table 2.

In the third round, this generated list was sent to all of the experts via email or was administered in the form of in-person or phone interviews. To determine content validity, these items were rated using a four-point Likert-type scale similar to the one described by Davis (1992) and often used in the literature. In this particular scale, 1 = not important, 2 = somewhat important, 3 = quite important, and 4 = highly important. The experts were asked to review each potential item in the list and rate the importance of each item in medical training using the four-point Likert-type scale. Experts were also requested to provide a justification for each rating.

In the fourth round, the same list of items were rated once again by all experts at least one week later to confirm their original rankings, and the experts were allowed to offer additional feedback as well as additional items to be included in the list of criteria. The provision of the opportunity to include additional criterion items is similar to a procedural step used in the study by Lisk *et al.* (2014). The experts used the identical procedure followed in round three for reliability purposes.

In addition, the experts were asked an additional five questions about their views of the overall Delphi method and its use in gathering experts' opinions in general, in regards to anatomy, and in regards to pelvic anatomy specifically. Initially, these questions were intended to be asked of all experts in a group setting as a focus group session. However, due to scheduling conflicts between each of the experts, these questions were included at the end of the second generated list distributed during the fourth round.

Data Analysis
During rounds three and four, those items which were rated as a three or a four by the majority of experts were retained in the instrument. Responses were used to calculate a content validity index (CVI) and determine whether an item would be retained. The CVI for each item, also known as the item-level content validity index (I-CVI) is the proportion of experts who rate the item as 3 or 4 (important) versus 1 or 2 (not important). As recommended in the literature (Davis 1992), items with I-CVI's that were less than 0.80 were dropped from the evaluation tool. The CVI for the entire set of items, also known as the scale-level content validity index (S-CVI), was also calculated by determining the mean proportion of items rated as a three or a four by the recruited experts. This particular form of S-CVI is also called the average calculation method (S-CVI/Ave). To maintain the rigor of the instrument, the minimum S-CVI/Ave for the instrument was established as 0.90 as recommended by the literature (Waltz *et al.* 2005, Polit and Beck 2006). Therefore, any S-CVI/Ave calculation below 0.90 would have warranted further analysis of rounds three and four.

For both the third and fourth rounds, the internal consistency via Cronbach's alpha was calculated for each item, for the entire scale, and for the experts' ratings between the third and fourth rounds using SPSS 20.0.

Results

Delphi Method: Rounds 1 and 2
Of the seven experts recruited in this study, three agreed to participate in round one via interviews. The pediatric urologist and the urogynecologist participated in the in-person interviews while the other urologist participated in a phone interview. The experts reviewed their transcribed responses in a second round and confirmed their responses with only minor corrections via email. The other four experts, who included a gynecologist and three basic science faculty members in anatomy, submitted their first-round responses via email.

From the responses of these seven experts, a list of 156 items was generated. The items included anatomical structures and features, 3D relationships among structures, and key views, orientations, and positions. Overall, the list contained 111 anatomical structures and features, 19 3D relationships between structures, and 26 key views, orientations, and positions. The anatomical structures and features included 33 skeletal structures and features, 14 muscular structures and features, 38 organs and other soft tissue structures (12 male-specific, 14 female-specific, 17 non-sex-specific structures), 13 vascular structures, 8 nerves, and 5 spaces (Table 2).

Delphi Method: Round 3
In round three, all 156 items were rated by six of the experts on a four-point Likert-type scale. The urologist dropped out of the study. Table 3 displays the ratings for each item by each of the experts along with its respective I-CVI. It also provides the S-CVI/Ave for the entire scale (0.79).

Delphi Method: Round 4
All 156 items from the third round were rated again by six of the experts. The only change made to the list was the addition of the individual abdominal wall muscles rather than simply grouping them as abdominal wall muscles, bringing the total number of criteria to 159. Each of these muscles was then

continued on next page

rated as a separate item. Table 4 displays the ratings for each item by each of the experts along with its respective I-CVI. It also provides the S-CVI/Ave for the entire scale (0.79).

Overall Content Validity
In the third and fourth rounds, the I-CVI values were in agreement for 136 of the 159 items, warranting either inclusion in or removal from the criteria list. For the remaining 23 items, the I-CVI values were not in agreement. These items had third- and fourth-round values whereby the former either warranted the removal of the items from the list or warranted the retention of the items in the list. Averages of these I-CVI values were calculated, providing I-CVI values that warranted the removal of 19 of the 23 items from the list of criteria. Four items were retained as their average I-CVI values were 0.83. Table 5 provides a list of those 23 items not in agreement between the third and fourth rounds, their third- and fourth-round I-CVI values, and their average I-CVI values. In addition to these 23 items, 39 more items were removed from the criteria list because they had I-CVI values below 0.80 in both the third and fourth rounds. Table 6 provides a list of all 62 items that were removed from the criteria list overall, their third- and fourth-round I-CVI values, and their average I-CVI values. Table 7 provides the final list of retained criteria with their third- and fourth-round I-CVI values and their average I-CVI values. After the removal of all 62 items, the final list of retained criteria contained a total of 97 items with an S-CVI/Ave value of 0.92. Figure 1 provides a graphical representation of the average I-CVI values for the items within each category of the final criteria list.

Internal Consistency
Coefficient Alpha supported rater consistency between rounds 3 and 4. These results are presented in Table 8. Criteria items that were retained in the reliability analysis met a certain threshold suggesting that they were consistent across items. All of the retained items had a Coefficient Alpha above 0.90, which according to George and Mallery (2003) indicates excellent internal consistency. These results are presented in Table 9. The Coefficient Alpha results in Table 10 showed two expert raters who rated items differently from the other four experts. One expert was a clinician, and one was a basic scientist. According to George and Mallery (2003), Coefficient Alpha values as low as 0.7 are acceptable while values between 0.70 and 0.60 are questionable. Although Nunnally (1967) stated that coefficient values of 0.60 or 0.50 were sufficient in "early stages of research," he increased the coefficient value to 0.70 in a later edition (Nunnally 1978).

Experts' Perceptions
Overall, participants found the Delphi method to be an effective technique for gathering experts' opinions. Although several experts found the multiple rounds to be repetitive, redundant, and tedious, all of the experts deemed it to be efficient. In addition, one expert in particular admitted to being unaware of another less time-consuming method to acquire experts' opinions.

Another expert made a fair point that the specific needs of the individual learning cohorts have considerable variation. These learning groups include graduate- and undergraduate-level medical, dental, kinesiology, and other allied health students as well as undergraduate anatomy and physiology students. This expert continued to say that "a robust program can be tailored to the group" based on the anatomical knowledge they are expected to learn. This specific study did not ask experts to provide pelvic criteria important specifically for certain cohorts of students, but it did ask the experts to consider criteria important for a 3D anatomical model of the pelvis, in general. However, experts were asked to provide criteria "most critical to portray" to anatomy students and trainees regarding 3D relationships among anatomical structures. These anatomy students include graduate- and undergraduate-level medical, dental, kinesiology, and other allied health students as well as undergraduate anatomy and physiology students.

In addition, during the first-round survey, experts were asked questions regarding their perception of the need for 3D anatomy in general anatomy education as well as the need for 3D anatomy in pelvic anatomy education, specifically for pre-health students, medical students, and medical residents. Overall, there were mixed responses from the experts in terms of their views on which form of 3D anatomy education was more or less valuable for these particular cohorts of students and trainees. However, the responses, according to Figure 2, suggest that two of the three basic science experts viewed general 3D anatomy to be more valuable for medical students and medical residents than pre-health students while two of the four clinical experts viewed general 3D anatomy to be more valuable for pre-health and medical students than medical residents. In turn, the responses, according to Figure 3, suggest that two of the three basic science experts and two of the four clinical experts viewed pelvic 3D anatomy to be useful for all three learning groups.

Discussion

Validity
Validity is an important concept to be considered in the development of standardized instruments and tools that will serve a purpose in teaching and training individuals in a particular field or discipline. In addition, the techniques used to validate these instruments and tools must also be validated using methodical, well-documented procedures. Therefore, measures were taken to validate each round of the Delphi process. Hasson and Keeney (2011) mention several different types of validity that were considered in the process of validating the Delphi method.

continued on next page

During round one as well as all three additional rounds of the Delphi method, the expert participants remained anonymous to each other in order to prevent experts from influencing the responses of one another, thus helping to maintain the criterion-related validity (Hasson and Keeney 2011). According to both Schmidt (1997) and Okoli and Pawlowski (2004), in the second round, experts should review their first-round responses as interpreted and organized by the investigator to approve or revise them as necessary in order to maintain the construct validity (Hasson and Keeney 2011) of the instrument being produced. Although this particular round was necessary only for the expert responses acquired via personal or in-phone interviews, it ensured that the criteria recorded from the experts during their first-round interviews were indeed the criteria they intended to profess to the interviewee.

Despite the fact that one expert dropped out of the study, a total of six experts were deemed sufficient because, according to Lynn (1986), I-CVI values as low as 0.78 are acceptable when there are six or more experts. In consideration of content validity, the average calculation method for scale content validity (S-CVI/Ave) was used because it is not quite as stringent as the universal agreement calculation method (S-CVI/UA) which is the proportion of items that received a rating of 3 or 4 from all of the experts. The average calculation method was used since the probability of reaching one hundred percent agreement among experts decreases as the number of consulted experts increases (Polit and Beck 2006). Since the S-CVI/Ave of the final list of criteria was greater than 0.90, the overall content of the list is rigorously valid (Waltz *et al.* 2005, Polit and Beck 2006).

A four-point Likert-type scale, instead of a traditional five-point Likert scale, was used to determine content validity in order to ensure that experts made a definitive decision about whether an item was or was not important. Similar four-point scales have been used in other studies (Waltz and Bausell 1981, Lynn 1986, Davis 1992). In fact, one study in particular argued the use of a four-point scale to deter experts from settling on irresolute ratings (Lynn 1986). In this particular study, the ultimate goal was to establish consensus among the experts involved as to which items are truly important to consider in pelvic models.

Reliability

This fourth administration of the Delphi method allowed for a test-retest measure for ascertaining whether the same group of experts can come to a consensus at a different time frame similar to a study performed by Uhl (1975) who found that consensus was obtained in three rounds of Delphi administration. Such consensus established over multiple time periods corroborates the test retest reliability of the Delphi process.

For this reason, items with an I-CVI less than 0.80 in the third round were not immediately removed from the overall list of criteria before the second rating as they were in the Lisk *et al.* study (2014). The experts were allowed to rerate all of the original criterion items again in the fourth round at least one week after the third round.

Internal Consistency

In the third round, each of the following component items (as they were numbered in Table 2) had zero variance, and they were removed from the scale in regards to reliability measurements: 1, 2, 3, 4, 6, 10, 11, 12, 13, 15, 16, 18, 22, 26, 57, 63, 78, 79, 85, 92, 148, 153, and 154. In the fourth round, each of the following component items (as they were numbered in Table 2) had zero variance, and they were removed from the scale in regards to reliability measurements: 1, 2, 3, 4, 6, 13, 14, 15, 16, 56, 57, 58, 62, 63, 72, 78, 79, 85, 92, 95, 96, 97, 144, 145, 153, 155, 170, 171, and 172. The determinant of the covariance matrix in both the third and fourth rounds was zero or approximately zero. Statistics based on its inverse matrix could not be computed, and they were displayed as system missing values.

Coefficient Alpha is a single measure of internal consistency. Just because an item may not be consistent with other items does not mean it cannot be considered. Coefficient Alpha can also be thought of as measuring unidimensionality. Perhaps those items that were inconsistent were just as important, but they referenced another dimension. The inconsistency between expert number two and expert number five could potentially be due to differing areas of expertise or different views on items that are most important. Further exploration would require a detailed factor analysis.

Experts' Perceptions

Overall, there were mixed perceptions of basic science and clinical faculty regarding the value of general and pelvic 3D anatomy for pre-health students, medical students, and medical residents according to Figure 2. The fact that 67% (two out of three) basic science experts viewed general 3D anatomy to be more valuable for medical students and residents than for pre-health students could have been because basic science faculty consider that pre-health students require less detail in general 3D anatomical knowledge in their earlier years of anatomy instruction. In fact, one of the faculty members with such a view felt that pre-health students needed less detail in general 3D anatomical knowledge than medical students and medical residents.

Fifty percent (two out of four) of the clinical experts might have viewed general 3D anatomy to be more valuable for pre-health and medical students than medical residents because these experts acknowledged the fact that pre-health students benefit from learning from models before learning

continued on next page

from real human anatomy. Moreover, one of the clinical experts stated that while 3D anatomy is initially important for medical residents' surgical training in the form of simulations, eventually they will be performing surgeries on real people and will have less need for 3D simulations. Nevertheless, one clinician emphasized the fact that training medical residents on surgical simulations before performing surgical procedures on real patients is critical to preventing injuries to patients.

On the other hand, the majority of experts viewed pelvic 3D anatomy to be important for pre-health students, medical students, and medical residents, according to Figure 3. The fact that that 67% of the basic science experts and 50% of the clinical experts viewed pelvic 3D anatomy to be useful for all three learning groups might have been because all of these experts understood that the pelvic region is an anatomical area of high complexity and limited accessibility. In fact, most of these experts stated this view explicitly in their responses. Since 3D pelvic models typically reduce the congestion of fascia and adipose tissue that are visible in cadavers, they can potentially reduce the cognitive load of students and trainees when they are learning the 3D relationships.

The Development of the Pelvic Criteria List

Removed items

Most of the anatomical structures that were removed from the initial generated list of pelvic criteria at the end of the third and fourth rounds collectively were soft tissue structures that cannot be readily viewed using computerized tomographic (CT) imaging and that have relatively little clinical importance. Work is already in progress by the authors in reconstructing a model of the male pelvis from CT scans using Amira® software. Thus, this valid list of criteria will help ensure that only those anatomical structures which are truly important for students and residents to learn will be included in completed pelvic models.

Most of the removed anatomical structures included ligaments, fascial tissue, relatively obscure organs, veins, and smaller, less significant nerves. However, some of the structures that were removed based on the overall experts' opinions were surprising. For instance, the Fallopian tubes (oviducts) as anatomical structures were ultimately removed from the list, but basic science and clinical experts in the third and fourth rounds asserted their importance in understanding their relationships to the ovaries, uterine wall, and vasculature and in understanding gynecological surgical landmarks, respectively. In addition, the obturator nerve was deemed important by basic science and clinical experts as it is a common source of iatrogenic injury due to misrecognition during surgical procedures. In fact, obturator nerves are typically at risk during lymphadenectomy for endoscopic radical prostatectomy procedures (Stolzenberg *et al.* 2016, Teber *et al.* 2009) which are very common urological surgeries. Nevertheless, an important consideration to keep in mind is the fact that not all anatomy learners intend to pursue clinical practice in a healthcare field. Many anatomy students throughout the world take anatomy and physiology, and the students within this particular cohort who have no intention of becoming physicians, dentists, or other allied health professionals may only need to know prominent anatomical structures that provide them with a basic context of an anatomical region as complex as the pelvis.

Moreover, some structures, despite their clinical importance, were removed because they are not easily distinguished on CT scans, they consist of different component structures by a another name, they are not important surgical landmarks in the experts' respective disciplines, or they can be easily viewed and appreciated in other learning formats such as traditional two-dimensional (2D) images from textbooks. For example, the piriformis muscle was ultimately removed because the gynecological expert claimed that it is poorly differentiated on CT scans although overall muscle groups can be visualized and because one of the urologists claimed that it was not an important urological surgical landmark. Nevertheless, all of the basic scientists asserted its importance; one of them even reasoned that it was critical for understanding 3D relationships. Clinically, knowing the location of the piriformis muscle in relation to the sciatic nerve is also important for understanding the connection between piriformis syndrome and sciatica (Hopayian *et al.* 2010).

Some of the structures, such as the perineal body, external urethral sphincter muscle, and abdominal wall muscles, were removed from the criteria list due to their low item-level content validity index (I-CVI) values (less than 0.80). However, all of the clinicians considered these structures to be important surgical landmarks or important urological surgical structures. Some basic science faculty might have viewed these structures with less importance probably because they do not encase the pelvic cavity, or they are difficult to distinguish in CT scans. Although these structures were important for urologists or gynecologists, not all anatomy students will become surgeons or urologists. The majority of anatomy students may only need a basic overview of anatomical structures within the pelvis region as suggested by some of the experts' comments from the first-round survey.

Finally, some items which were removed from the criteria list, despite their clinical importance, will still be potentially present in a pelvis model constructed from CT scans. For instance, the five spaces that have surgical relevance might be visualized as much less attenuated regions in a pelvis model if the boundaries created by the surrounding soft tissue structures are also visualized.

Retained Items

Interestingly, several trends in terms of what basic scientists viewed as important as opposed to what clinicians viewed as important arose from the experts' ratings. One major trend

continued on next page

that arose from the data in Tables 3 and 4 suggested that those items that clinicians, but not basic scientists, rated highly were important surgical structures or landmarks. Another trend from these tables seemed to suggest that those items which basic scientists, but not clinicians, rated highly were important for teaching anatomy students the 3D relationships of anatomical structures to each other in the pelvic region. On average, however, both basic science and clinical experts, according to Figure 1, tended to rate most of the skeletal structures and features as well as most of the organs and other soft tissues structures as important.

The Importance of Virtual 3D Anatomy

Historically, cadaveric dissection has been the traditional method by which students have learned human anatomy. One study in particular has shown that cadaveric dissection is the best way for students to learn the 3D relationships of anatomical structures (Wright 2012, Bergman *et al.* 2015). The use of prosections is also growing in popularity and effectiveness in conveying these 3D concepts to students (Nnodim 1990, Samarakoon *et al.* 2016). However, cadaveric dissections and prosections are not always as effective in helping students understand the 3D relationships of anatomical structures in regions of high complexity and limited accessibility.

Since some undergraduate institutions, especially, do not have access to cadavers due to finances or ethical matters (Robbins *et al.* 2009, Lempp 2005), there is a need for research to confirm what types of 3D technology and which specific virtual 3D models are efficient to use in the instruction of anatomy students. Therefore, methodical procedures for valid and reliable instrument construction for virtual 3D model assessment need to be followed to ensure the creation of valid virtual 3D models of anatomical regions such as the pelvis.

Conclusions

Overall, a valid and reliable list of pelvic criteria was successfully created using a multiple-round Delphi method procedure. Although the removal of some items from the criteria list was surprising, one expert suggested that virtual 3D anatomical models can be customized for different groups of anatomy learners. Moreover, all of the experts viewed the Delphi method process to be an effective way of gathering experts' opinions.

Limitations

First of all, participants in a multiple-round Delphi method study might have found the repetitious or redundant nature of the iterative rounds and questions to be annoying. In addition, participants might have also considered the multiple rounds to be time consuming and even unnecessary, especially the third and fourth rounds since they accomplished the same goal of establishing consensus among experts. However, the duplication of these two rounds was necessary to ensure test-retest reliability of the instrument. Nevertheless, these first two limitations might have been regarded as sufficient reason for participants to discontinue their participation in the study, thus compelling participants to drop out. In fact, in this study, one of the experts dropped out, thus hindering the overall content validity of the criteria list. However, the I-CVI/Ave was still calculated at a rigorously valid value.

Moreover, experts who submitted their responses electronically for both the third and fourth rounds could have easily duplicated or copied their responses from previous iterations of the instrument. Such actions could have potentially biased the validity of the overall instrument. However, in this study, no ratings between the two rounds for any of the experts were completely identical. In fact, several experts submitted their third- and fourth-round responses in different formats.

Furthermore, while the Delphi method was useful for collecting experts' opinions about anatomy in general, soliciting opinions from experts regarding complex regions of anatomy such as the pelvis was more problematic. Securing adequate numbers of experts proved to be a challenge, given the varied locations of experts, their varying time availability to commit to participation, and their potential need for incentives. Fortunately, a range of three to ten experts was deemed sufficient for a Delphi method procedure (Lynn 1986). Plus, ensuring that participants were willing to complete all phases of the survey proved to be difficult, especially since experts were located at a distance from the study director.

Future Directions

Future directions will include the use of this valid list of criteria to inform the addition of anatomical structures to an existing virtual 3D model of the male pelvis and to validate the male pelvis model once it is completed. Additional future studies will focus on the impact of the validated virtual 3D male pelvis model on learning and retention in anatomical studies. This list of criteria could also be used by researchers to validate any virtual 3D anatomical pelvis model that already exists. By using the criteria list to validate current pelvis models, researchers can ensure that the models are accurate before implementing them in student instruction. Further exploration might also involve modifying validated models to tailor them to specific target groups of learners. Modifications could be made by consulting experts who have a basic understanding of the target learners' baseline knowledge at the beginning of their respective educational program, who have routine interactions with the learners, and who have experience teaching or training them.

continued on next page

Acknowledgements

Sincere gratitude is extended to Charles R. Pound, MD, Urology Division Chief at the UMMC; Christopher M. Bean, a UMMC pediatric urologist; Paul Moore, MD, a UMMC urogynecologist; and G. Rodney Meeks, MD, Professor Emeritus of Obstetrics and Gynecology from the UMMC; all of whom served as clinical experts in pelvic anatomy during the Delphi method process. Additional thanks are extended to Ranjan Batra, PhD; Susan Warren, PhD; and Yuefung Lu, PhD; all of whom served as basic science experts in pelvic anatomy from the Department of Neurobiology and Anatomical Sciences at the UMMC during the Delphi method process. Much gratitude is extended to Andrew Notebaert, PhD, Director of the Clinical Anatomy Program at UMMC, Marianne L. Conway, MD, Assistant Professor of Neurobiology and Anatomical Sciences, and Susan Warren, PhD, Professor of Neurobiology and Anatomical Sciences at the UMMC, for their contributions to the review of the first-round survey before its dissemination to study participants. Much appreciation is also given to Andrew Notebaert, PhD, and Erin Dehon, PhD, Director of Faculty Development and Behavioral Science Education and Associate Professor of Emergency Medicine at UMMC, for their contributions to reviewing the format of the generated criteria lists during the third and fourth rounds of the Delphi method. And finally, special thanks is extended to Kenneth Thompson, PhD, the registrar in the Office of Institutional Research and the Office of the Registrar at Millsaps College in Jackson, Mississippi, who served as the statistician in this research study.

About the Authors

Edgar R. Meyer, MAT**,** graduated from Millsaps College with a BS in Biology and a BA in Classical Studies, and from Delta State University with an MA in Teaching. He is currently a fourth-year PhD graduate student in the Clinical Anatomy Graduate Program in the Department of Neurobiology and Anatomical Sciences at the University of Mississippi Medical Center (UMMC) in Jackson, Mississippi. He has served in formal assistant teaching roles in the medical gross anatomy, medical histology and cell biology, and medical neuroscience and behavior courses. He has also served in voluntary assistant teaching roles in the medical developmental anatomy, dental developmental anatomy and histology, and dental gross anatomy courses at the institution. His research interests include the evaluation of the impact of stereoscopic models of both pelvic and middle and inner ear anatomy on first-year medical students' short- and long-term retention of the knowledge of 3D relationships between structures.

Amber James, MCA**,** earned here BA in biochemistry from the University of Mississippi and her MS in clinical anatomy from the UMMC. She is currently a third-year medical student at the UMMC. She has served as a teacher's assistant in medical gross anatomy and medical neuroscience and behavior and as a voluntary assistant for the pre-matriculation gross anatomy course for incoming medical students. Her research interests include the formulation of a pelvic model that can serve as a teaching tool for both the learning of basic anatomy and relevant anatomical relationships in radiological imaging.

Dongmei Cui, PhD, MD (hon.), is an Associate Professor of the Department of Neurobiology and Anatomical Sciences, Clinical Anatomy Division, at the UMMC. She is investigating and developing digital teaching tools in anatomy education, and she serves as the mentor and Principal Investigator for students' work in the 3D Virtual Anatomy laboratory.

Literature cited

Azer SA, Azer S (2016) 3D anatomy models and impact on learning: A review of the quality of the literature. *Health Professions Education*. 2: 80-98. doi: 10.1016/j.hpe.2016.05.002

Bergman EM, de Bruin ABH, Vorstenbosch MATM (2015) Effects of learning content in context on knowledge acquisition and recall: A pretest-posttest control group design. *BMC Med. Educ.* 15: 133. doi: 10.1186/s12909-015-0416-0

Beyersdorff D, Schiemann T, Taupitz M, Kooijman H, Hamm B, Nicholas V (2001) Sectional depiction of the pelvic floor by CT, MR imaging and sheet plastination: Computer-aided correlation and 3D model. *Eur. Radiol.* 11: 659-664. doi: 10.1007/s003300000561

Brown PM, Hamilton NM, Denison AR (2012) A novel 3D stereoscopic anatomy tutorial. *Clin. Teach.* 9: 50-53. doi: 10.1111/j.1743-498X.2011.00488.x

Caves R (1988) Consultative methods of extracting expert knowledge about clinical competence. In: R. Ellis (Ed.) Professional Competence and Quality Assurance in the Caring Professions (pp. 199-299). London, UK: Chapman Hall.

Cui D, Wilson TD, Rockhold RW, Lehman MN, Lynch JC (2017) Evaluation of the effectiveness of 3D vascular stereoscopic models in anatomy instruction for first year medical students. *Anat. Sci Educ.* 10: 34-45. doi: 10.1002/ase.1626

Dalkey NC, Helmer O (1963) An experimental application of the Delphi method to the use of experts. *Management Science.* 9: 458-467. doi: 10.1287/mnsc.9.3.458

Davis LL (1992) Instrument review: Getting the most from a panel of experts. *Applied Nursing Research.* 5: 194-197. doi: 10.1016/S0897-1897(05)80008-4

George D, Mallery P (2003) SPSS for Windows Step by Step: A Simple Guide and Reference. 11.0 Update (Fourth Edition). Boston, MA: Allyn & Bacon.

Hasson F, Keeney S (2011) Enhancing rigour in the Delphi technique research. *Technological Forecasting & Social Change.* 78: 1695-1704. doi: 10.1016/j.techfore.2011.04.005

continued on next page

Helmer O (1967) Systematic Use of Expert Opinions (Document Number P-3721). Santa Monica, CA: The RAND Corporation.

Hisley KC, Anderson LD, Smith SE, Kavic SM, Tracy JK (2008) Coupled physical and digital cadaver dissection followed by a visual test protocol provides insights into the nature of anatomical knowledge and its evaluation. *Anat. Sci. Educ.* 1: 27-40. doi: 10.1002/ase.4

Hopayian K, Song F, Riera R, Sambandan S (2010) The clinical features of the piriformis syndrome: A systematic review. *Eur. Spine J.* 19:2095-2109. doi: 10.1007/s00586-010-1504-9

Hoyek N, Collet C, Di Rienzo F, De Almeida M, Guillot A (2014) Effectiveness of three-dimensional digital animation in teaching human anatomy in authentic classroom context. *Anat. Sci. Educ.* 7: 430-437. doi: 10.1002/ase.1446

Khot Z, Quinlan K, Norman GR, Wainman B (2013) The relative effectiveness of computer-based and traditional resources for education in anatomy. *Anat. Sci. Educ.* 6: 211-215. doi: 10.1002/ase.1355

Lawshe CH (1975) A quantitative approach to content validity. *Personnel Psychology.* 28: 563-575. doi: 10.1111/j.1744-6570.1975.tb01393.x

Lempp HK (2005) Perceptions of dissection by students in one medical school: Beyond learning about anatomy. A qualitative study. *Anatomy Teacher*. 39: 318-325. doi: 10.1111/j.1365-2929.2005.02095.x

Lisk K, Flannery JF, Loh EY, Richardson D, Agur AMR, Woods NN (2014) Determination of clinically relevant content for a musculoskeletal anatomy curriculum for physical medicine and rehabilitation residents. *Anat. Sci. Educ.* 7: 135-143. doi: 10.1002/ase.1393

Lynn MR (1986) Determination and quantification of content validity. *Nursing Research*. 35: 382-386. doi: 10.1097/00006199-198611000-00017

Maggio MP, Hariton-Gross K, Gluch J (2012) The use of independent, interactive media for education in dental morphology. *J. Dent. Educ.* 76: 1497-1511.

Marsh KR, Griffin BF, Lowrie Jr. DJ (2008) Medical student retention of embryonic development: Impact of the dimensions added by multimedia tutorials. *Anat. Sci. Educ.* 1: 252-257. doi: 10.1002/ase.56

Müller-Stich BP, Löb N, Wald D, Bruckner T, Hans-Peter M, Kadmon M, Büchler MW, Fischer L (2013) Regular three-dimensional presentations improve in the identification of surgical liver anatomy—A randomized study. *Anat. Sci. Educ. BMC Med. Educ.* 13: 131. doi: 10.1186/1472-6920-13-131

Nicholson DT, Chalk C, Funnell WR, Daniel SJ (2006) Can virtual reality improve anatomy education? A randomized controlled study of a computer-generated three-dimensional anatomical ear model. *Med. Educ.* 40: 1081-1087. doi: 10.1111/j.1365-2929.2006.02611.x

Nnodim JO (1990) Learning human anatomy: By dissection or from prosections? *Med. Educ.* 24: 389-395. doi: 10.1111/j.1365-2923.1990.tb02456.x

Nunnally JC (1967) Psychometric Theory. New York, NY: McGraw-Hill.

Nunnally JC (1978) Psychometric Theory (Second Edition). New York, NY: McGraw-Hill.

Okoli C, Pawlowski SD (2004) The Delphi method as a research tool: An example, design considerations, and applications. *Information & Management*. 42: 15-29. doi: 10.1016/j.im.2003.11.002

Parikh M, Rasmussen M, Brubaker L, Salomon C, Sakamoto K, Evenhouse R, Ai Z, Damaser MS (2004) Three dimensional virtual reality model of the normal female pelvic floor. *Ann. Biomed. Eng.* 32: 292-296. doi: 10.1023/B:ABME.0000012749.79488.d6

Polit DF, Beck CT (2006) The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health.* 29: 489-497. doi: 10.1002/nur.20147

Qayumi AK, Kurihara Y, Imai M, Pachev G, Seo H, Hoshino Y, Cheifetz R, Matsuura K, Momoi M, Saleem M, Lara-Guerra H, Miki Y, Kariya Y (2004) Comparison of computer-assisted instruction (CAI) versus traditional textbook methods for training in abdominal examination (Japanese experience). *Med. Educ.* 38: 1080-1088. doi: 10.1111/j.1365-2929.2004.01957.x

Robbins BD, Tomaka A, Patterson CIJ, Styn G (2009) Lessons from the dead: The experiences of undergraduates working with cadavers. *OMEGA – Journal of Death and Dying.* 58: 177-192. doi: 10.2190/OM.58.3.b

Samarakoon LB, Vithoosan S, Kokulan S, Dissanayake MM, Anthony DJ, Dissanayake V, Jayasekara R (2016) Anatomy of teaching anatomy: Do prosected cross sections improve students understanding of spatial and radiological anatomy? *Anatomy Research International.* Article ID 8984704, 4 pages. doi: 10.1155/2016/8984704

Sergovich A, Johnson M, Wilson TD (2010) Explorable three-dimensional digital model of the female pelvis, pelvic contents, and perineum for anatomical education. *Anat. Sci. Educ.* 3: 127-133. doi: 10.1002/ase.135.

Schmidt RC (1997) Managing Delphi surveys using nonparametric statistical techniques. *Decision Sciences.* 28: 763-774. doi: 10.1111/j.1540-5915.1997.tb01330.x

continued on next page

Stolzenburg JU, Rabenatt R, Do M, Lee B, Truss MC, McNeil A, Burchardt M, Jonas U, Liatsikos E (2016) Complications of endoscopic extraperitoneal radical prostatectomy (EERPE): Prevention and management. *World J. Urol.* 24: 668-675. doi: 10.1007/s00345-006-0133-8

Teber D, Gozen AS, Cresswell J, Canda AE, Yencilek F, Rassweiler J (2009) Prevention and management of ureteral injuries occurring during laparascopic radical prostatectomy: The heilbrohm experience and a review of the literature. *World J. Urol.* 27: 613-618. doi: 10.1007/s00345-009-0428-7

Uhl NP (1975) Consensus and the Delphi process. Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D. C., April 1975) (ERIC Document) ED 104201.

Waltz CF, Bausell RB (1981) Nursing Research: Design, Statistics, and Computer Analysis. Philadelphia, PA: F. A. Davis.

Waltz CF, Strickland OL, Lenz ER (2005) Measurement in Nursing and Health Research (Third Edition). New York, NY: Springer Publishing Co.

Womble MD (1999) Anatomy and computers: A new twist to teaching the oldest medical course. *Bioscene*. 25: 15-17.

Wright SJ (2012) Student perceptions of an upper-level, undergraduate human anatomy laboratory course without cadavers. *Anat. Sci. Educ.* 5: 146-157. doi: 10.1002/ase.1265

■

Back to TOC