

Developing a Tool to Assess Physical Therapist Educational Program Quality With Engagement Theory: The American Council of Academic Physical Therapy Benchmarks for Excellence Task Force

Amy E. Heath, PT, DPT, PhD, Peter Altenburger, PT, PhD, Jacklyn H. Brechter, PT, DPT, PhD, Gary Chleboun, PT, PhD, Diane U. Jette, PT, DSc, FAPTA, Gary Pike, PhD, Denise Schilling, PT, PhD, Kimberly Topp, PT, PhD, FAAA, and Barbara Tschoepe, PT, DPT, PhD

Background and Purpose. The American Council of Academic Physical Therapy (ACAPT) convened the Benchmarks for

Amy Heath is an associate professor of practice and chair in the Department of Physical Therapy at Simmons College, 300 The Fenway, Boston, MA 02115 (amy.heath@simmons.edu). Please address all correspondence to Amy Heath.

Peter Altenburger is an associate professor and chair in the Department of Physical Therapy at Indiana University.

Jacklyn H. Brechter is an associate professor in the Department of Physical Therapy at Chapman University.

Gary Chleboun is a professor and director in the School of Rehabilitation and Communication Sciences at the Ohio University.

Diane U. Jette is a professor and associate chair in the Department of Physical Therapy at the MGH Institute of Health Professions.

Gary Pike is a professor in the School of Education at the Indiana University.

Denise Schilling is a professor and associate dean in the College of Allied Health Professions at the Western University of Health Sciences.

Kimberly Topp is a professor and chair in the Department of Physical Therapy and Rehabilitation Science at the University of California San Francisco.

Barbara Tschoepe is a visiting professor and director in the Doctor of Physical Therapy program at the University of Vermont.

The authors report no conflicts of interest.

Received January 29, 2018, and accepted March 2, 2018.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.aptaeducation.org).

Copyright © 2018 Academy of Physical Therapy Education, APTA

DOI: 10.1097/JTE.0000000000000048

Excellence (BenEx) Task Force with a charge to define and assess excellence in physical therapist education. The purpose of this article is to describe the process employed by the BenEx Task Force in the development of a tool to measure physical therapist education program excellence, to provide evidence for the validity and reliability of the tool, and to describe the future goals of the BenEx Task Force.

Method/Model Description and Evaluation.

The BenEx Task Force members adopted the Engagement Theory of Program Quality as a framework for defining excellence. In 2013, the task force developed the initial *Physical Therapist Measure of Educational Program Quality* survey. The task force worked closely with a Web site design company to develop a comprehensive item database, survey format, survey delivery mechanism, data-entry process, and data summary and interactive display platform. The final student survey included 36 elements representing 11 attributes within the 5 clusters of quality hypothesized by the Engagement Theory, and the faculty survey included 38 elements representing 13 attributes within the 5 clusters.

Outcomes. In 2015, 88 of 193 (46%) ACAPT-eligible programs participated in the survey: 706 students, 717 faculty, and 88 program directors. The analyses revealed some areas of the survey that may require revision; however, 35 of 36 elements on the student survey and 30 of 38 elements on the faculty survey met the expected acceptable alpha reliability coefficient levels ($\geq .70$) of internal consistency.

Discussion and Conclusion. Defining excellence in physical therapist education is a difficult undertaking. The process used by the BenEx Task Force to begin the work of defining excellence in physical therapist education underscores the need for

continuous engagement of stakeholders. Enlisting literature from other areas of education and a theoretical framework helped to present a cohesive structure that allowed ACAPT members to agree to the process enough to move the project forward. Furthermore, a transparent process promoted success. Continued data collection and analysis will determine the inclusion/exclusion of specific items that may not fit well into particular elements, any necessary survey revisions, and facilitation of the development of benchmarks in physical therapist education.

Key Words: Benchmark, Engagement, Program quality, Education, Physical therapist.

BACKGROUND AND PURPOSE

The American Council of Academic Physical Therapy (ACAPT) was formally approved by the American Physical Therapy Association (APTA) House of Delegates in 2013 with a vision of being “the leading voice to promote, achieve, and sustain excellence in academic physical therapy” (<http://www.acapt.org/about/who-we-are>). Based on this vision, a primary goal for this new organization was to develop mechanisms to assist academic physical therapist programs in achieving excellence. As a necessary step toward this goal, a task force was convened with a charge to define and assess excellence in academic physical therapy. To this end, the 2011 Education Leadership Conference (ELC) focused on identifying the key elements of educational excellence. With nearly 200 attendees participating, the principles of appreciative inquiry¹ were used to guide small group discussions about the elements of excellence in physical therapist education prompted by the following statement: “If we want to flourish and excel, we need to understand and

articulate the ends that are most important to us, then do all we can to excel at the activities necessary to reach them.” The discussions were recorded and then summarized by a work group of 10 members. Data from these summaries were used to derive a definition of excellence, identify vital components of excellence, and identify mechanisms to achieve excellence. The summary of the findings was posted to the ACAPT community’s Sharepoint site for further discussion and input by members in January 2012 (Appendix 1, Supplemental Digital Content 1, <http://links.lww.com/JOPT/A21>). Subsequently, the Benchmarks for Excellence (BenEx) Task Force of ACAPT proceeded to use the data and feedback to 1) operationalize the definition of excellence; 2) develop a tool to measure the construct; and 3) recommend a benchmarking process for programs to work toward achieving excellence.

Benchmarking is a systematic approach to measurement, comparison, and evaluation that focuses on finding and implementing best practices in any industry or enterprise.² As an important mechanism for achieving best practices, benchmarking is similar to a continuous quality improvement process in which one’s current status is evaluated using one or more criteria of importance, and the desired state is envisioned. Goals are determined by how well the current state matches the desired state, and a plan is developed to achieve the desired goals.

Achievement of excellence in physical therapist education is a process driven by self-assessment and programs’ desire for ongoing improvement in the areas of excellence that they deem relevant to their mission. Through a benchmarking process, physical therapist educational programs may match their own performance on identified criteria for excellence against all or, a relevant sample of, other programs and use that information to determine areas of strength and those areas needing improvement. The benchmarking process includes assessment from diverse perspectives, allowing for triangulation of information. For example, physical therapist educational programs might seek the perspectives of students, faculty, and administrators. The benefits of benchmarking for a program might include 1) enhancing the learning environment for faculty and students; 2) improving student and faculty outcomes; and 3) demonstrating accountability to various stakeholders including students, institutional administration, potential donors, accreditation commissions, employers, and, ultimately, patients. As envisioned by the founding members of ACAPT, the process of benchmarking could enhance programs’ ability to move beyond meeting

minimal standards of achievement toward the goal of excellence.

The purpose of this article is to describe the process employed by the BenEx Task Force in the development of a tool 1) to measure physical therapist education program excellence, 2) to provide evidence for the validity and reliability of the tool, and 3) to describe the future goals of the BenEx Task Force.

METHOD DESCRIPTION AND EVALUATION

Benchmarking

The BenEx Task Force initially sought to understand the process of benchmarking, as this understanding seemed essential to developing sound recommendations to ACAPT members. In the academic setting, benchmarking is a quality assurance and enhancement practice.³ In 1997, Higgs and McMeeken⁴ published an article describing a 5-year benchmarking process involving all of the physiotherapy programs in Australia and New Zealand. The primary steps in this process included the following: 1) identifying indicators of excellence, 2) defining gold standards (benchmarks) relative to the indicators, 3) developing mechanisms for collecting data, 4) collecting relevant data, 5) sharing data among participating programs, and 6) implementing follow-up actions. Jones et al⁵ suggested that quality assurance and benchmarking in dental education required stakeholders to recognize their needs and develop objectives based on those needs. Achieving these objectives then becomes “quality.” The authors recommended four essential process elements for attaining quality: 1) clarity of goals and objectives, 2) clarity of methods for evaluation, 3) a system to make improvements to the methods, and 4) a review process to assess system changes. The literature clearly establishes fundamental processes for benchmarking: a need for identifying indicators of excellence and a mechanism for capturing performance related to these indicators.

Engagement Theory of Program Quality

During its early work, the BenEx Task Force members conducted a literature search focusing on excellence in higher education. This search and subsequent deliberations led to the adoption of the Engagement Theory of Program Quality as a framework for defining excellence as a first step in determining potential benchmarks. The Engagement Theory of Program Quality was first proposed by Haworth and Conrad⁶ in 1997 based on a qualitative study of national scope. The investigators interviewed nearly 800 stakeholders from approximately 50 master’s

degree programs across 11 fields of study, with the goal of identifying the characteristics of high-quality programs. High-quality programs were defined as those that gathered input from all stakeholders to “create enriching learning experiences for their students that positively affect their growth and development.”⁶ (p.15) Based on their findings, the investigators posited 5 clusters of program quality: 1) diverse and engaged participants and leaders; 2) participatory cultures; 3) interactive teaching and learning; 4) connected program requirements; and 5) adequate resources. Several attributes of quality are included in each of the 5 clusters (Appendix 2, Supplemental Digital Content 2, <http://links.lww.com/JOPT/A22>).

The Engagement Theory was subsequently tested among a variety of higher education programs and constituents. Those included master’s and doctoral degree students and faculty in one university department of educational administration,⁷ students and faculty in several education programs in member institutions of the Council of Christian Colleges and Universities,⁸ students from a sample of 48 US master’s in counseling programs,⁹ and a sample of approximately 2500 master’s and doctoral degree students at one university.¹⁰ The Engagement Theory has also been applied in qualitative studies designed to understand the characteristics of successful interdisciplinary graduate programs¹¹ and to explore the quality in distance education programs.¹² Thus, the theory has demonstrated fairly wide applicability and relevance to graduate-level education.

DEVELOPMENT PROCESS

Applying the Engagement Theory of Program Quality

Following the literature review, discussions, and agreement to adopt the Engagement Theory of Program Quality, the BenEx Task Force began to match the theoretical framework to the components of excellence in physical therapist education derived from the work of ACAPT members completed in 2011–2012. Appendix 2 (Supplemental Digital Content 2, <http://links.lww.com/JOPT/A22>) includes the clusters of the theoretical framework in the first column and the second column includes the attributes of excellence identified by members of ACAPT. In addition, the task force reviewed a study by Grignon et al¹³ examining the outcomes statements of physical therapy graduates from a national sample of programs. In this study, all physical therapist education programs were asked to submit the statements of expected graduate outcomes that they used to meet the Commission on Accreditation in Physical Therapy Education (CAPTE) criterion that assesses the

alignment of program goals and mission.¹⁴ Seventy-five programs responded, and a qualitative conventional content analysis identified core concepts in the statements. The BenEx Task Force matched those core concepts to the clusters and attributes of the Engagement Theory (the third column of Appendix 2, Supplemental Digital Content 2, <http://links.lww.com/JOPT/A22>). The task force then identified quantitative program characteristics that might be considered to be reflective of each of the attributes suggested by the Engagement Theory (the fourth column of Appendix 2). Finally, the task force developed qualitative measures that they believed reflected the excellence in physical therapist education programs based on the theoretical framework (the fifth column of Appendix 2).

Development of Survey Items

In 2013, the task force developed the initial *Physical Therapist Measure of Educational Program Quality* survey using all of the components listed in Appendix 2: clusters and attributes identified by the Engagement Theory, components of excellence in physical therapy education derived from members at ELC, 2011, themes from the program outcomes study,¹³ and quantitative and qualitative characteristics identified by the task force. Survey items were then developed to reflect each component (Supplemental Digital Content 2, <http://links.lww.com/JOPT/A22>). Task force members engaged in rigorous discussion regarding the meaningfulness, usefulness, and measurability of each item. Items were reviewed for clarity and consistency and to determine if and how each item linked to the original theoretical elements. Following this rigorous review, items were modified, added, or omitted. Items were placed into three separate surveys, one for each stakeholder: administrators, faculty, and students. The three surveys were then aligned so that responses from different stakeholders to similar items could be used to inform the underlying construct. For example, both faculty and student surveys included an item related to the impact of out-of-class activities on students' professional development. Asking the same question of different stakeholders allowed results to be triangulated. This was seen as particularly important for items related to qualitative constructs because multiple data points could provide evidence for validity of the responses and capture potential difference in perspectives of various stakeholders.

The survey was piloted by nine task force members' programs: 42 faculty, 4 administrators, and 103 students. Using the pilot results, individual survey items were examined to determine if the variance in responses

was sufficiently high to warrant retaining the items; survey items were edited to enhance clarity and consistency of formatting, to decrease the time to complete the survey, and to improve users' experience with the survey. Feedback was also solicited from the participants regarding questions, concerns, or suggestions for improvement. Because of the pilot data and feedback, the administrator survey was combined with the faculty survey in order to prevent administrators from inadvertently not completing the faculty survey.

The revised surveys were then beta-tested by a larger sample of program volunteers. Task force members solicited beta-test participation from 56 programs. Of those invited, 36 programs participated in the first beta-test: 135 students, 150 faculty, and 25 administrators.

Data from the pilot and the first beta-test were combined and the aggregate results, as well as feedback about the process, were shared with attendees at the 2015 Combined Sections Meeting (CSM) and 2015 Education Leadership Conference (ELC). At the ELC preconference Open Forum and as a report to the ACAPT Board, the task force demonstrated the online platform for reporting the survey results. At various meetings during both conferences, the task force requested feedback from assorted stakeholders. Based on this feedback and the available data, a second beta-test was conducted in the fall of 2015. All member institutions of ACAPT were invited to participate.

Development of Survey Deployment Mechanism

The task force worked closely with OpenArc (Pittsburg, PA), a Web site design company, to develop a comprehensive item database, survey format, survey delivery mechanism, data-entry process, and data summary and interactive display platform. Once a program representative agreed to participate, the program director was provided with information on how to upload names and email addresses of faculty and students. OpenArc then sent emails inviting these individuals to participate. Each email contained a personalized login, which enabled responses to be anonymously and securely entered and stored. Reminder emails were sent to individuals who failed to login or complete the survey by a designated date. Students were asked to complete the survey within 3 months of completion of their program. The task force continued to gather feedback regarding the survey process and initiated improvements when feasible.

Final Survey Characteristics

The final student survey included 36 elements representing 11 attributes within the 5 clusters

of quality hypothesized by the Engagement Theory. The faculty survey included 38 elements representing 13 attributes within the 5 clusters. The difference in number of attributes in the student versus faculty survey was due to each survey including some uncategorized attributes. Uncategorized attributes were items that addressed elements outside of the Engagement Theory. For example, the student survey includes two uncategorized attributes: "program performance" and "exemplary programs" compared to the one uncategorized attribute "overall program quality" for the faculty survey. Additionally, three attributes, "diverse and engaged faculty," "residency," and "support for faculty" were only included for the faculty survey. The average reported time to complete either survey was 32 minutes, and all data were temporally confined (by year) in order to enable year-to-year comparisons.

Survey Testing

At the conclusion of the second beta-test period, a total of 193 programs had been invited to participate. A total of 88 programs participated (46%). This total included 706 students, 717 faculty, and 88 program directors. The results from the second beta-test for student and faculty surveys are presented in this article.

Separate analyses were conducted for the student and faculty surveys. Survey items were combined into hypothesized elements within each of the attributes identified by the Engagement Theory. Interelement correlations and Cronbach's *alpha* were used to assess the internal consistency of each element.¹⁵ As Tavakol and Dennick¹⁶ noted, *alpha* reliability is widely used in medical education studies to assess internal consistency reliability. As a rule of thumb, an *alpha* reliability coefficient of .70 or greater was taken as an indication of acceptable internal consistency. As an additional check on the construct validity of the attributes, correlations among elements thought to be related to the same attribute were calculated and examined. This approach allowed judgments to be made about the structural validity of the attributes and elements.¹⁷ It was expected that the correlations among elements related to the same attribute would be moderate to high ($r \geq .40$).¹⁸

OUTCOMES

Student Survey Results

Cronbach's *alpha* ranged from .51 to .93 for the various elements. Correlations among elements ranged from $-.30$ to $.78$. Table 1 presents the *alpha* reliability coefficients and

Table 1. Cronbach's Alpha and Interelement Correlations for the Student Survey

Cluster	Attribute	Element	Alpha	Interelement Correlations				
Diverse and engaged participants	Engaged learners	Leadership development (1 item)		1				
		Program involvement (3 items)	0.81	0.48	1			
		APTA member (1 item)		0.07	0.09	1		
Participatory cultures	Shared program direction	Program engagement (5 items)	0.86	1				
		Program beliefs (4 items)	0.89	0.59	1			
		Community of learners						
			In-class activities (8 items)	0.77	1			
			Out-of-class activities (9 items)	0.87	0.74	1		
			DPT core faculty (8 items)	0.88	0.46	0.45	1	
			DPT associated faculty (7 items)	0.84	0.44	0.42	0.63	1
		Risk taking/interactive learning	Confidence (8 items)	0.91	1			
	Faculty create environment (4 items)		0.88	0.60	1			
Interactive teaching and learning	Mentoring/cooperative learning	Engagement core faculty (7 items)	0.85	1				
		Opportunity to participate (3 items)	0.78	0.41	1			
		Formal assessment (5 items)	0.70	0.52	0.33	1		
		Experiential learning	Attend local conferences (1 item)		1			
			Attend state conferences (1 item)		0.36	1		
			Attend national conferences (1 item)		0.18	0.34	1	
			Attend international conferences (1 item)		-0.01	0.22	0.1	1
		Impact professional development (9 items)	0.86	-0.18	-0.2	-0.3	-0.02	1
Connected program requirements	Depth/breadth coursework	Learning strategies (5 items)	0.76	1				
		Contributions to development (5 items)	0.82	0.56	1			
		Additional learning strategies (4 items)	0.77	0.39	0.62	1		
		Curricular features (7 items)	0.89	0.48	0.62	0.56	1	
		Tangible products	Capstone experiences (5 items)	0.51	1			
			Capstone contribution (5 items)	0.89	0.40	1		
			Capstone encouraged you (4 items)	0.93	0.31	0.73	1	
			Clinical experience (6 items)	0.84	0.28	0.51	0.54	1
		Clinical contribution (6 items)	0.92	0.23	0.59	0.60	0.65	1
Adequate resources	Infrastructure	Program resources (9 items)	0.89	1				
		Rating of services (6 items)	0.91	0.57	1			

Table 1. Cronbach's Alpha and Interelement Correlations for the Student Survey continued

Cluster	Attribute	Element	Alpha	Interelement Correlations	
Uncategorized		Academic/clinical resources (7 items)	0.90	0.71	0.55
		Quality academic faculty (3 items)	0.81	0.57	0.4
	Program performance	0.81	1		
	Exemplary programs	Areas of distinction (4 items)	0.92	0.78	1
		Exemplary graduates (10 items)	0.83	1	
		Program success (10 items)	0.88	0.57	1

Abbreviation: APTA = American Physical Therapy Association. Bolded items are below the accepted levels of .70 for internal consistency and .40 for interelement correlation.

correlations among elements organized by attribute for the student survey. For those elements based on a single survey item, no reliability coefficients are provided.

With one exception, the elements' *alpha* reliability coefficients were greater than or equal to .70. The exception, "Your DPT program includes the following culminating/capstone experiences" [capstone experiences], produced a reliability coefficient of .51. An examination of the interelement correlations reveals that most correlations exceeded the .40 threshold. The correlations between the two uncategorized elements representing exemplary program performance were also greater than the .40 threshold.

Interelement correlations for three attributes, "engaged learners," "experiential learning," and "tangible products" were frequently below .40. If the single item concerning whether the student was a member of APTA [APTA member] was excluded, the remaining two elements comprising the "engaged learners" attribute were moderately correlated (.48). None of the elements comprising the "experiential learning" attribute produced correlations in excess of .40, and several of the correlations were negative. Three of the correlations for the elements associated with "tangible products" were less than the .40 threshold; however, these low correlations appear to be a product of unreliability of measurement for one element. If the "culminating/capstone experience" element [capstone experiences] that produced the low reliability coefficient mentioned previously was not included, all of the remaining interelement correlations were greater than .40.

Faculty Survey

Cronbach's *alpha* ranged from .36 to .94 for the various elements. Correlations among elements within attributes ranged from .18 to .76. Table 2 presents the *alpha* reliability coefficients and correlations among elements organized by attribute for the faculty survey.

Measures of internal consistency for responses to the faculty survey were generally lower than the corresponding measures for the student survey. Because unreliability of measurement tends to attenuate the magnitude of correlations, it is not surprising that the interelement correlations were somewhat lower for the faculty survey as well.

In total, eight elements had estimates of internal consistency that were less than the .70 threshold. In some cases (eg, "Over the past 3 years, students had the opportunity to participate." [opportunity to participate] and "You believe students would rate your use of the following teaching and learning strategies as meaningful to their development." [how

students rate strategies]), the magnitudes of the correlations were only slightly lower than the .70 threshold. In others ("As a DPT faculty member, you create an environment that encourages students to:" [encourage students to] and "Your DPT curriculum requires students to complete the following culminating experiences:" [capstone experiences]), *alpha* reliability coefficients were extremely low.

An examination of the interelement correlations in Table 2 reveals that convergence of the correlations was observed for five attributes, "diverse and engaged faculty," "engaged leaders," "professional residency," "support for faculty," and "infrastructure." An additional five attributes would produce acceptable interelement correlations ($r \geq .40$) if suitably modified. For example, the elements related to "shared program direction" evidenced an acceptable correlation if the element "You are involved in the following:" [general involvement] was dropped from the construct. Likewise, dividing the four elements comprising the "community of learners" attribute into two distinct areas, namely, 1) "... you engaged students in the following in-class activities" [in-class activities]/"... you engaged students in the following out-of-class activities" [out-of-class activities] and 2) "... you believe DPT core faculty collectively were" [core faculty]/"... you believe DPT associated faculty were" [associated faculty], produced acceptable interelement correlations.

DISCUSSION AND CONCLUSION

Defining excellence in physical therapist education is a difficult undertaking. Previous researchers⁷⁻¹⁰ have evaluated a variety of educational programs using survey items, initially identified and tested by Mustan,⁷ that were shown to represent the clusters proposed by the Engagement Theory. These studies supported the construct validity of the 5 clusters represented by Engagement Theory. The BenEx Task Force used the work of these investigators along with the work generated by ACAPT members to develop survey items specific to physical therapist education.

The process used by the BenEx Task Force to begin the work of defining excellence in physical therapist education underscores the need for continuous engagement of stakeholders at every stage of development. The stages included work by a small task force of ACAPT members who sought insight from the existing literature, held multiple discussions within and outside the group, engaged with stakeholders by seeking participation and feedback, and amended and adapted the process along the way. In addition, the assistance of a professional Web site

Table 2. Cronbach's Alpha and Interelement Correlations for the Faculty Survey

Cluster	Attribute	Element	Alpha	Interelement Correlations				
Diverse and engaged participants	Diverse and engaged faculty	Interaction with diversity (4 items)	0.94	1				
		Success with diversity (10 items)	0.84	0.47	1			
		Administrator involvement (19 items)	0.87	1				
		Student opportunities (4 items)	0.73	0.41	1			
Participatory cultures	Shared program direction	General involvement (8 items)	0.73	1				
		Program involvement (16 items)	0.88	0.17	1			
		Three-year involvement (8 items)	0.85	0.20	0.76	1		
	Community of learners	In-class activities (8 items)	0.71	1				
		Out-of-class activities (11 items)	0.80	0.68	1			
		Core faculty (7 items)	0.80	0.24	0.22	1		
		Associated faculty (7 items)	0.89	0.18	0.18	0.56	1	
		Risk taking/interactive learning	Last 3 years (3 items)	0.79	1			
		Graduates able to (8 items)	0.86	0.32	1			
		Encourage students to (4 items)	0.45	0.24	0.45	1		
Interactive teaching and learning	Mentoring/peer learning	Engage with students (7 items)	0.75	1				
		Opportunity to participate (3 items)	0.67	0.35	1			
		Formal assessment (5 items)	0.64	0.26	0.24	1		
	Experiential learning	Impact professional development (6 items)	0.79	1				
		Impact additional activities (3 items)	0.75	0.46	1			
		Explicit learning objectives (4 items)	0.74	0.21	0.25	1		
Connected program requirements	Depth/breadth coursework	Teaching/learning strategies (5 items)	0.58	1				
		How students rate strategies (5 items)	0.68	0.53	1			
		Rate additional strategies (4 items)	0.59	0.28	0.46	1		
		Curricular contribution (7 items)	0.83	0.21	0.35	0.29	1	
	Professional residency	Student experiences (3 items)	0.67	1				
		Impact of experiences (3 items)	0.84	0.57	1			

Table 2. Cronbach's Alpha and Interelement Correlations for the Faculty Survey continued

Cluster	Attribute	Element	Alpha	Interelement Correlations	
	Tangible products	Capstone experiences (5 items)	0.36	1	
		Capstone encouragements (4 items)	0.87	0.25	1
		Quality of experiences (5 items)	0.80	0.50	1
	Support for faculty	Scholarly resources (4 items)	0.80	1	
		Faculty policies (5 items)	0.76	0.46	1
Adequate resources	Infrastructure	Program resources (9 items)	0.86	1	
		Services for faculty (10 items)	0.88	0.59	1
		Academic/clinical resources (7 items)	0.84	0.58	1
Uncategorized	Overall program quality	Quality of faculty (3 items)	0.79	1	
		Quality student experience (3 items)	0.75	0.69	1
		Areas of distinction (4 items)	0.71	0.23	1
		I believe DPT program (4 items)	0.77	0.59	0.68
					1

Bolded items are below the accepted levels of .70 for internal consistency and .40 for interelement correlation.

developer and statistician was instrumental in allowing smooth distribution of surveys and gathering and analyzing data. It seems likely that as programs come to understand the power of the online platform to visualize and interact with the data, greater participation will follow.

Development Process

As with any organization the size of ACAPT, we encountered differing opinions and agendas among stakeholders. Enlisting literature from other areas of education and a theoretical framework helped to present a cohesive structure that allowed ACAPT members to agree to the process enough to move the project forward. Furthermore, a task force comprising representatives from various types of institutions (private/public; urban/rural; teaching/research intensive) and a transparent process with consistent updates and opportunities to voice concerns, ask questions, and provide feedback promoted the support from multiple stakeholders needed for success.

Initially, the logistical challenges included Web site development and the actualization of the requests of the task force in a user-friendly format. Once the survey was constructed and administered to the initial pilot and beta-test participants, it became clear that communication with participants regarding the survey was essential, and the logistics of managing the communication were challenging. For example, the program director email list generated from the CAPTE database was not completely accurate, as program directors had changed or were on sabbatical, or there was an error in the transcription of the email address. Such logistic challenges required a considerable outreach effort from the task force to correct email addresses, address misunderstandings, answer questions regarding the process, and offer apologies for confusion and inconveniences. Because of this particular challenge, the administrator and faculty surveys were consolidated, and clearer directions were developed.

Survey release dates were also challenging. Faculty and administrators reported that students were often completing a final clinical experience when the survey was released, resulting in a poor response rate because faculty and program directors had a difficult time explaining to students the significance of the survey and why it was important to complete. Faculty surveys were often released at hectic times in the semester and became a low priority for completion. In the future, these issues will be addressed with "windows" for completion—a 3-month scheduled time frame that will be communicated to the

student and faculty well before the dates. This change, along with reminder emails, may increase future response rates.

Another area of challenge was stakeholder inclusion. The task force sought input and consultation from all interested parties not only as a means of transparency but also to ensure that there were no elements of excellence that were overlooked. In this attempt to be inclusive, the task force collaborated with the ACAPT Research-Intensive Programs in Physical Therapy (RIPPT) consortium to address how scholarly activity relates to excellence in physical therapist education. This collaboration with the RIPPT consortium resulted in other consortia feeling excluded. Because exclusion is antithetical to the pursuit of defining excellence and the execution of excellence, the task force redoubled its efforts to get input from as many members of ACAPT as possible.

Survey Testing

For the most part, measurement characteristics of the survey met the expected acceptable levels of internal consistency. For the student survey (35 of 36) and faculty survey (30 of 38), elements met the .70 alpha coefficient threshold. Further analyses revealed, however, some areas of the survey that may require revision. For example, the element related to culminating/capstone experiences showed poor internal consistency. This may be due to the fact that many of the capstone experiences listed in the item set are mutually exclusive (eg, “faculty advised individual research project/presentation” and “faculty advised group research project/presentation”). Similarly, the fact that none of the elements comprising the “experiential learning” attribute produced correlations in excess of .40 and had several negative correlations suggests that the elements represent stand-alone measures rather than a unified attribute. Examples of “experiential learning” elements include the following: attendance at local, state, national, and international conferences and impact on professional development. In spite of the fact that elements did not fit into a unifying attribute, they may provide useful information for program evaluation and ongoing data collection and analysis of the survey will be used to refine and improve survey quality. Any future survey revisions will likely minimize the number of elements with low alpha coefficients and interelement correlations.

Application of Survey Results

One significant remaining challenge is determining how survey results will be used.

From the initial conception of this project, some ACAPT members were concerned that the outcomes would be used to compare one program to another in order to recruit students and faculty. From the beginning, the task force tried to address this concern. Currently, this concern is being addressed in the following ways: 1) data are only reported to CAPTE recorded program directors (or their designee as communicated to the Task Force) in the aggregate, eliminating the ability of programs to attribute any specific findings to any one student or faculty member and 2) with the use of filters in the outcomes platform, individual program results can only be compared to a minimum of three other programs. If, by selecting a filter, there are fewer than three programs to compare against, the program director will not be able to garner results from that analysis. These features are consistent with the purpose of benchmarking as a continuous quality improvement process to allow programs to reach their goals in concert with their mission. This purpose will continue to be the primary intention moving forward as benchmarks are established, with the secondary intention of elevating program quality toward excellence.

Another remaining challenge is survey fatigue. Program directors, students, and faculty are asked to participate in many surveys including the CAPTE Annual Accreditation Report (AAR), surveys supporting dissertations and other student projects and likely others. The current annual data collection process has been implemented to increase awareness and participation and to allow for enough data to accurately assess survey performance. The task force will be recommending that each program complete the survey once every 2 years to decrease survey fatigue and increase response rates while keeping the relevant information up-to-date.

Limitations

Every effort was made to obtain a representative sample of students from DPT programs. Nevertheless, there may be sampling bias. Conclusions about the construct validity of the surveys are also limited by the focus on the internal consistency of the attributes and elements. As Messick¹⁷ observed, the convergence of measures with external constructs is another important component of construct validity, and future research should examine the relationships between clusters and elements, elements of the student and faculty surveys, and other measures of DPT program quality. At this time, however, there does not appear to be another measure against which to make this type of assessment.

Future Directions

Based on the results of the analyses of the structure of the surveys, discussions will ensue about the inclusion of items that may not fit well into particular elements. One important future step to reduce redundant data entry and survey fatigue is to combine the *Physical Therapist Educational Program Quality* survey with the AAR or to arrange for information from the AAR to autopopulate in the *Physical Therapist Educational Program Quality* administrator survey. The BenEx Task Force plans to use the results from the 2016 implementation to further analyze the hypothesized elements of the Engagement Theory of Program Quality and publish the results. Final adaptation of the survey will ensue, and another year of data were collected in 2017.

While data collection continues with the *Physical Therapist Educational Program Quality* survey, the BenEx Task Force will be working to determine benchmarking best practices or models to facilitate this process. The methods used by the task force will include the same transparency and opportunities for feedback as were employed during the previous stages. The process will inform the final phase of the project, establishing benchmarks for excellence in physical therapist education. The hope is that once benchmarks are developed, programs will engage in a systematic approach to measurement and evaluation that focuses on quality improvement methods to push beyond meeting minimal standards to achieving excellence.

REFERENCES

1. Busche G. The appreciative inquiry model. In: Kessler EH, ed. *Encyclopedia of Management Theory*. Thousand Oaks, CA: Sage Publications; 2013.
2. Gift RG. Benchmarking. In: Gift RG, Kinney CF, eds. *Today's Management Methods: A Guide for the Health Care Executive*. Hoboken, NJ: Wiley, John & Sons; 1996:245-261.
3. Hobson R, Rolland S, Rotgans J, et al. Quality assurance, benchmarking, assessment and mutual international recognition of qualifications. *Eur J Dent Educ*. 2008;12:92-100.
4. Higgs J, McMeekin J. Achieving quality in physiotherapy programmes through benchmarking. *NZ J Physiother*. 1997;25:19-22.
5. Jones M, Hobson R, Plasschaert A, et al. Quality assurance and benchmarking: An approach for European dental schools. *Eur J Dent Educ*. 2007;11:137-143.
6. Haworth JG, Conrad C. *Emblems of Quality in Higher Education: Developing and Sustaining High-quality Programs*. Needham, MA: Allyn and Bacon; 1997.
7. Mustan T. *Operationalization and Preliminary Testing of the Engagement Theory*.

- Madison, WI: University of Wisconsin-Madison; 1998.
8. Kornelis PC. *Faculty Members' and Students' Perceptions of Quality in Master of Education Programs within Member Schools of the CCCU*. Vermillion, SD, University of South Dakota; 2004.
 9. Warden S. Testing the engagement theory of program quality in CACREP-accredited counselor education programs. *Counselor Edu Supervision*. 2012;51:127-140.
 10. Wang HM. *What Matters in Graduate School? Exploring Patterns of Student Engagement, Academic and Personal Development*. Tampa, FL: Association for Institutional Research; 2003.
 11. Newswander LK, Borrego M. Engagement in two interdisciplinary graduate programs. *Higher Edu*. 2009;58:551-562.
 12. Carranza SL. *A Grounded Theory of High-quality Distance Education Programs: Student Perspectives*. Charleston, SC: BiblioBaazar; 2008.
 13. Grignon TP, Henley E, Lee KM, Abentroth MJ, Jette DU. Expected graduate outcomes in US physical therapist education programs: A qualitative study. *J Phys Ther Educ*. 2014;28:48.
 14. Standards and required elements for accreditation of physical therapist education programs. 2016.
 15. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951; 16:297-334.
 16. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ*. 2011;2:53.
 17. Messick S. Validity. In: Linn RL, ed. *Educational Measurement*. 3rd ed. New York, NY: American Council on Education Series on Higher Education and Macmillian; 1993: 13-103.
 18. BMJ. 11. Correlation and regression. <http://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/11-correlation-and-regression>. Accessed June 29, 2017.