USA Program Description

Overview
The Ultrasound Student Assessment (USA) Program was created in response to a request by ultrasound instructors to find an improved method of selecting students for ultrasound education. Collectively, many instructors believed that the existing standardized tests were inadequate at identifying students with skills well suited to performing ultrasound studies. This is problematic in that most schools have a limited number of admission slots. Each seat filled by an unqualified student results in an expense for both the student and the school that does not have ROI potential, leading to disgruntled faculty, students and administrations. Unlike the TEAS V (formerly known as the HOBET) and other entrance exams used by schools as part of their entrance requirements, the USA program has been created to test specific skills required to learn and perform ultrasound clinical studies. The USA program is taken as a blind admission test. As with other exams used in assessing candidates for college or college programs, it is anticipated that the results of the USA program will be used in conjunction with other admission criteria.

Program Details

History
In 2008 at the SDMS annual conference, a group of instructors were discussing the difficulties associated with students who, nearing the end of their program, were incapable of developing the skills necessary to adequately perform ultrasound studies. Frustrated, the instructors asked if Pegasus Lectures could develop a method that would improve the selection process so that the strengths associated with the ability to become a proficient ultrasound professional could be determined before accepting students into an ultrasound program. Their belief was that if a method could be created, both the school and the students would benefit. The school would not invest so much time and money into a student only to find close to the end of the program that the student did not possess the required skills. Additionally, the student would not find themselves close to the end of the program either dropping out or failing out, having already spent considerable time and financial resources, with no potential return on investment. The instructors also believed that better identifying more adept students would reduce the number of disgruntled students and relieve stress on themselves and their programs, especially as many programs accept so few students.

Structure of the Research Project
As a result of this conversation, a three phase, longitudinal research project was commenced. The continuation of the project at each phase was contingent upon “success” in the prior phase. The results of each prior stage impacted the direction of each subsequent phase.
**Phase I:**
Phase I was designed to determine if there was sufficient reason to conduct the ongoing research as well as to determine direction, should a Phase II be warranted. Specifically, a survey of ultrasound instructors was sent out to determine:

- what percentage of instructors believed there was a problem that needed solving
- what methods instructors were using to accept students
- what qualities/skills they believed best correlated with successful students
- what qualities/lack of skills they believed best correlated with unsuccessful students
- what percentage of their students dropped/failed out of their program

The specifics of Phase I were presented at the SDMS annual conference in October of 2009. Following is a summary of the most pertinent points:

- 75 programs responded
- Instructors indicated that they accepted about 27% of the applicants to their programs (this point mattered because if there was no selectivity (i.e. 100% acceptance rate) there would be no value in an entrance exam
- About 21% of the students selected do not graduate on to perform ultrasound studies (graduation rate of 79%)
- Entrance requirements were far from being standardized
- Interview information:
  - 50% of the programs perform interviews
  - less than 16% test problem solving and only 20% ask technical questions
  - only 57% responded that they believed their selection process to be effective
  - academic issues was the top reason students dropped out
- Asked a series of questions and then performed a statistical analysis to determine if a Phase II was warranted, and to serve as a guideline for creating Phase II

Realizing that surveys are imperfect tools, there were two methods used to analyze the results gathered. For the first method, I analyzed what instructors believed to be true using descriptive statistics. For the second method, I compared stated admission policies and instructor beliefs with a metric of “success.” This metric was graduation rates from each program. It is important to note that this is not a perfect metric as schools can have very different standards, so higher standards may give the impression of a worse admission policy. However, even given this limitation, we deemed this the best possible metric for Phase I as it was at least “reasonable” and not subjective.

**Summary of Instructor Beliefs**

**Academic success associated with**
- Motivation
- Study habits

**Ability to scan associated with**
- Coordination
- Spatial recognition (abilities)
- Dedication/persistence

**Operate Effectively in Hospital/Lab setting associated with**
- Communication
- Social Skills/People person
- Integrity and compassion
Summary of Correlations with Graduation Rates

1. Strong positive correlation with selectivity
2. For requirements, high GPA had the strongest correlation (followed by recommendations and evaluations)
3. Requiring only a high school diploma (no college courses or other degrees) had a significant negative correlation
4. Strong correlation with performing interviews for selection
5. Interviews which tested problem solving (highest) and technical questions (2nd) had strong positive correlations
6. Instructors who ranked problem solving high in priority had the strongest positive correlation with graduation rates (note that problem solving shows up both from fact and from opinion)
7. Instructors who ranked work experience high in priority had the weakest correlation with graduation rates

Specific Results of Phase I:

- Although interviews had one of the best correlation with school success, many schools stated that it was not practical/possible because of number of candidates or, more often, because of state restrictions.
- There were multiple characteristics which correlated with success both based on metric comparison and from opinion based responses.
- The characteristics which came up repeatedly were:
  - Motivation/Persistence
  - Problem solving
  - Coordination/Dexterity
  - Spatial ability
  - Communication/integrity (decision making)
- When asked why students could not learn to scan well, the responses were:
  - Spatial limitations (50.00%)
  - Coordination issues (32.43%)
  - Drive/lack of persistence (9.46%)
  - Other (17.57 %)
- Good Practices: (Positive Correlation)
  - High Selectivity (no surprise there)
  - Interviews
  - GPA requirement
  - Evaluations
  - Other (17.57 %)
- “Bad” Practices: (Negative Correlation)
  - High school diploma only
  - Entrance exam

Interestingly, there was a negative correlation with entrance exams and graduation rates. This result would seemingly obviate the need to continue the research. However, for the following reasons, it was believed that this conclusion was false because:

- Schools which offered an exam, (from the statistics) seemingly relied too heavily on the exam and not enough on other practices
The entrance exams were likely not geared (or ideally geared) toward what really matters in terms of ultrasound student selection (as evidenced by the initial discussion which prompted this study).

There were clear indications that there are many characteristics, if tested for, could help determine a priori who would more likely not succeed.

The conclusion of Phase I was that although not definitive, it was worth expending the effort to create a Phase II which gathered actual data related to the beliefs and characteristics that statistically seemed to correlate with more successful programs.

**Phase II**

In Phase II, a software platform was created to deliver an exam written specifically to test characteristics identified in Phase I. The test was created to address what was indicated in Phase I to be most likely important.

The exam was created to specifically test examinees for:

- **Coordination**
  - Software was created to specifically challenge a person to use both hands simultaneously while performing two different tasks simultaneously.

- **Persistence**
  - No one question or type of question tested persistence. Rather, persistence was gauged on multiple different parameters assessed throughout the program.

- **Problem solving**
  - Problem solving had a series of questions which require creative problem solving, not raw math abilities.

- **Spatial ability**
  - Spatial abilities were tested through pattern recognition and the ability to picture shapes created by the intersection of a plane and 3-D objects.

- **Judgment**
  - Judgment was tested through a series of situational questions in which a problem existed and the examinee must respond how they would address the problem.

- **Graph interpretation**
  - Questions based on reading, interpreting, and extrapolating information from graphs were given.

- **Visualization**
  - Noisy images (with speckle noise presentation similar to that of ultrasound) were given. The noise level was varied and the noise level recorded when the examinee could identify the object of the image.

- **Logic**
  - Questions were given that test various forms of logical fallacies, non sequiturs, and illogical extrapolation.

In addition to testing, data was collected on various topics to determine if there is any correlation between performance and certain activities. Data was collected regarding the following areas:

- **Years of education**
- **Subject taught**
- **Years teaching**
- Scan/not scan ultrasound
- Age
- Left or Right handed
- Gender
- Play an instrument (this parameter originated from Phase I as a write-in)

**Internet link to access the test**
- Restricted users to taking only once
- Did not allow user to stop and start
- Did not allow users to go back in exam (intentionally)
- Results were sent automatically to a server
- Results stored into “statistical spreadsheet”
- Calculations and correlations performed

A correlation calculation was then run on each personal characteristics with each topic tested. For those instructors who scan ultrasound, the correlations were strongest with dexterity, persistence, problem solving, and graph interpretation (in that order). For those instructors tested, playing an instrument had the strongest correlations with graph interpretation, judgement, dexterity, and completing the exam in less time.

**Conclusions from Phase II:**
- The parameters indicated from Phase I and tested in Phase II produced some “strong” correlations that for the most part made sense.
- Correlations indicated that there were parameters that vary with people who scan versus those who do not – this implies that doing Phase 3 made sense – although one could argue that these may be learned traits.
- The correlations were often stronger than expected – which was a very good sign that this approach (with some tweaks) could be useful.

**Phase III**

**Design:**
The design of Phase III required extensive software platform development. The goal of Phase III was to collect student test results as well as information regarding the quality of the student in actual ultrasound programs. This presented an interesting challenge. Pegasus needed to see all data while preserving anonymity of the students and instructors involved. Conversely, instructors needed to be able to add specific information regarding student aptitude without being able to access the student’s actual results from Phase III test (to avoid the introduction of bias).

Specifically, Phase III design included:
1. Adding additional graph questions
2. Adding more problem solving questions
3. Adding the ability to choose left or right hand mode for dexterity testing
4. Tweaking one problem solving question that all participants answered incorrectly
5. Removing teacher related information collection and replace with student related information collection
6. Creating anonymous data collection system for Pegasus analysis and calculation but still allow the data to be traceable by the instructor
7. Building a structure to capture student data from instructors and link to test results
8. Signing up schools to participate in Phase 3

Logistics of Phase III:
1. Students take exams at entrance to program (but no results are shown to instructors)
2. Instructors accept and reject students using “standard” process
3. Instructors periodically fill in information regarding accepted students such as grades in
   specific classes, ability to scan, scores on actual credentialing exams, rank in class, if the
   student dropped out and for what reasons, etc.
4. Analysis was then performed to see how well exam predicted who should have been admitted
   and those who should not have been admitted
5. Analysis comparing results of test with student performance metric
6. Modify exam to strengthen correlations (emphasize testing which best correlated with student
   success)

Analysis:
• Compile data from database
• Look at descriptive statistics
• Look for correlations and behaviors
• Set a metric to determine “good” vs. “bad”
• Create an algorithm using test data to create “score”
• Compare score against metric to determine quality of algorithm
• Change weighting functions to optimize algorithm

Some Descriptive Statistic Results of Phase III:
● 380 students took the tests
● 201 were identified as accepted
● The scores on actual credentialing exams were entered for 81 students
● 49 of the 201 known to be accepted were dropped
● Of the 49, 36 (about 75% of those dropped) dropped because of the inability to complete the
   clinical component of the coursework, the inability to scan adequately, inability to complete
   the physics coursework, behavioral issues related to judgement.
● Eleven students dropped the course due to family and health issue, two students dropped
   because of finances (these situations clearly cannot be identified by the test)
● Students who were ranked by the instructors as above average had higher dexterity scores
   than those students who were ranked below average and those students who dropped out

Creating a metric:
In order to calculate a statistical measure of “goodness,” a metric was needed against which to
compare results of the test algorithm. Creating a metric implies razor edge criteria, in essence using
the test algorithm counter to its intended purpose (please see “Discussion” below). The metric was
based on the direct feedback entered by the instructors, including if the student was dropped and
for what reason, ranking of the student on specific skills (on a scale from 1 (lowest) to 5 (highest)),
and actual scores on the ARDMS credentialing exam. The following outline the specific criteria
used to determine a “qualified” ultrasound student from an “unqualified” ultrasound student.
Qualified or High Cut:
- Physics
  - Score above 555 on the ARDMS credentialing exam
  - Average physics score as ranked by instructor above 4.0 (on a scale of 1 to 5)
- Specialty
  - Score above 555 on the ARDMS credentialing exam
  - Average specialty score as ranked by the instructor above 4.5 (on a scale of 1 to 5)

Unqualified or Low Cut:
- Physics
  - Score below 555 on the ARDMS credentialing exam
  - Average physics score as ranked by instructor below 3.0 (on a scale of 1 to 5)
  - Dropped from the program because of inability to complete or pass physics
- Specialty
  - Score below 555 on the ARDMS credentialing exam
  - Average specialty score as ranked by the instructor below 3.5 (on a scale of 1 to 5)
  - Dropped from the program because of inability to complete or pass specialty component
  - Dropped from the program because of inability to scan adequately

For student to be considered qualified, they had to be ranked as qualified in both physics and in a specialty.

Test Comparisons:
Students who achieved a positive score on the USA test were, for the purpose of the razor edge test, considered “qualified.” Students who achieved a negative score were considered unqualified. By comparing the results of the USA test versus the results of the metric, a simple calculation is possible to determine the sensitivity and overall accuracy of the test. The results are as follows:

Sensitivity: 83%
Accuracy: 75%

Discussion:
As mentioned previously, this calculation is the result of treating the result of the test as a razor’s edge in which a fraction of a point in either direction can mean that someone is considered as “qualified” or “unqualified,” violating the design and intent of the exam. The exam results are expected to be considered as part of a larger picture regarding student qualification, with the realization that the scores are part of a continuum and do not represent an absolute, specific threshold. Additionally, using rankings from different instructors for razor edge decisions as the metric creates another potential issue. Looking at mean rankings of students, it became very clear that some instructors were much more generous than other instructors when ranking their students. This was easily verified by comparing the average rank of the student relative to his/her ARDMS score. Some instructors ranked all of the students who passed the ARDMS exam with high scores below a 4 whereas other instructors ranked students with similar ARDMS scores as at or near 5.
When these two factors were taken into account and the calculation redone, the results were:

**Sensitivity:** 89%
**Accuracy:** 86%

**Conclusion:**
The USA program was created to test specific skills needed to be a successful ultrasound student and clinician. The test categories, type of questions, and algorithms were all developed using statistical measures that resulted from a several year longitudinal study. The test is intended to help inform school programs regarding the potential strength and weakness of each candidate as well as give the ability to compare candidates on a relative basis. No test can guarantee that all students who perform well on the entrance exam will perform well in the program, nor can it guarantee that students who do not perform well cannot become very good sonographers. However, the statistical results of comparing, in a blind fashion, student test results against instructor ranking and actual credentialing scores shows that the test has a very high correlation between scoring well and being ranked highly. Additionally, the test will continue to improve over time as new student data is added. Periodic reviews and statistical analysis will be performed to continue to update the exam content and adapt the algorithm to further improve the predictive ability of the USA program.

**Contact Information:**

972-564-3056
college@pegasuslectures.com