The ARRT Sonography Eligibility Pilot Study

Technical Report

October 2005

Table of Contents

1. **Introduction** 1
    Background 1
    Rationale for Certification in General Sonography 1
2. **Study Design** 2
    Sample 2
    Instrumentation 3
    Establishing a Score Scale 3
        Passing Score Study
        Total Scaled Scores
        Section Scaled Scores
3. **Results** 4
    Reliability and Validity 4
        Score Reliability
        Validity: Section Score Correlations
        Validity: Verifying Student Motivation
    Student Results 7
        Total Scores
        Section Scores
        Student Score Reports
    Program-Level Results 9
        Total Scores
        Section Scores
        Program Director Reports
4. **Summary and Conclusions** 11
5. **Appendices** 12
    A. Summary of Standard-Setting Study
    B. Sample Candidate (Student) Report
    C. Sample Program Director Reports

1. Introduction

Background

The ARRT has offered post-primary certification in sonography since January 2000. That program was intended for individuals certified in radiography who had gone on to train and work in ultrasound. As a post-primary certification program, eligibility was gained primarily through on-the-job experience. Candidates for certification were required to document that they had completed more than 1,500 sonographic examinations covering numerous anatomical regions.

In recent years, there has been an increase in the number of formal educational programs in sonography and a decreased reliance on on-the-job training. The profession is presently increasing its emphasis on formal education in other ways. For example, in 2005 a committee was formed to establish a national curriculum for sonography. In response to this new emphasis, and changes in demographics, the ARRT decided to also offer sonography as a *primary* examination – that is, a certification program available to non-RTs who are graduates of an accredited sonography program.

ARRT's primary certification in sonography is a practice-based exam that is consistent with the material taught by educational programs with a concentration in *general* sonography. The exam consists of 270 questions (250 scored plus 20 pilots) covering patient care, physical principles of ultrasound, and sonographic procedures involving abdomen, pelvis, gynecologic structures, obstetrics, superficial structures, and noncardiac chest. Exam results provide educators with a comprehensive evaluation of student outcomes, while the certificate affords hospitals and clinics with documentation that a sonographer is qualified to perform exams not just in a single clinical specialty, but in multiple areas.

Once the decision was made to modify the focus of the certification program, it became apparent that a pilot study should be conducted to verify that the exam would function as intended with the population of primary candidates. In addition, the results of a pilot study may be useful for verifying ARRT's assumption regarding the utility of a general exam for entry-level sonographers. If the data suggest that a single exam is not warranted, then there may be reason to consider separating the exam into multiple specialty exams.

The remainder of this report summarizes the results of the pilot study. The next section discusses ARRT's reason for offering an entry-level examination in general sonography. Subsequent sections describe the design of the pilot study and summarize the findings.

Rationale for Certification in General Sonography

Certification programs in sonography have, by convention, focused on specific areas of practice corresponding to numerous ultrasound specialties. It is common for someone to complete education or training in sonography, pass a physics exam plus one other exam in a clinical specialty (e.g., abdomen), and then designate themselves as a certified sonographer. One reason that the "specialty-based" model of certification has been so popular is that skills in sonography were traditionally acquired through clinic-based, on-the-job training. This usually meant that individuals acquired skills specific to the exams they performed in their clinical setting, and a certification mechanism that corresponds to specialty training has served the profession quite well.

As the ultrasound profession evolves, it has become evident that the specialty-based model of certification no longer serves the best interests of the public or profession. Many sonographers perform procedures that cross many specialties, even though they have passed an examination documenting their proficiency in just one clinical specialty. In contrast, anecdotal reports and empirical evidence indicate that sonographers work in multiple clinical areas. For example:

- A practice analysis completed by ARRT in 1999 indicated that general practice was quite prevalent. A factor analysis of a task analysis survey completed by 935 sonographers confirmed that practice could be carved into the domains of general sonography, vascular sonography, and echocardiography. Those

working in general sonography indicated they performed scans of the abdomen, pelvis, gynecologic structures, and superficial structures, and also performed obstetric exams.

- Two recent practice analyses completed by ARDMS demonstrated the same trend as ARRT data. A 2000 survey completed by 552 sonographers certified in abdomen indicated that, on average, this group spent 31% of their time also performing sonograms in obstetrics and gynecology. The average time spent performing exams of the superficial structures was 10%. A 2004 survey reported that over one-half of those certified in obstetrics and gynecology spent greater than 25% of their time performing abdominal exams, while over one-fifth of the group spent greater than 25% of their time doing breast exams.
- The composition of entry-level sonographers has changed and will continue to change. Historically, many sonographers came from among the ranks of R.T.s. Now an estimated 80% of new sonographers are not R.T.s, but gain entry into the field of medical imaging directly through sonography.
- Most formal educational programs offer a concentration in one or more of the three major areas of sonography: general, vascular, and echocardiography. The programs with a general concentration – and the students who graduate from them – will benefit from an exam that reflects the material covered by the curriculum.
- As health-care budgets continue to be trimmed, the need for multi-skilled sonographers will grow. An exam that covers multiple specialties can help ensure that the work force is broadly proficient.

The sonography profession has matured to the point of benefiting from the type of certification model utilized in fields such as medicine, law, nursing, and teaching. In most professions it is common to first become credentialed to engage in entry-level practice by passing a broad-based comprehensive exam. Then, after individuals gain experience, they may elect to seek certification in one or more specialties. This two-stage model of certification – a required general certificate followed by elective specialty certificates – has served other professions well for many years. ARRT certification in sonography is founded on such a model.

2. Study Design

Sample

In March 2005 ARRT contacted approximately 70 sonography educational programs about possible participation in the pilot study. Criteria for inclusion in the study included:

(a) The post-secondary institution in which the educational program is housed must be accredited by the degree-granting branch of a Regional Institutional Accrediting Agency. This meant that most programs were located in community colleges.[1]

(b) Students must satisfy all eligibility requirements and have graduated prior to taking the examination. Eligibility criteria included, among other things, completion of ARRT's clinical competency requirements. To satisfy the competency requirements, students must demonstrate to their program director or designee that they can competently perform the various sonographic procedures (e.g., spleen, kidney, thyroid, obstetrics) that span general sonography.

These requirements are consistent with the eligibility criteria for all ARRT primary examinations. A total of 115 students signed up to participate in the pilot study representing 18 accredited programs from all regions of the country. Programs were located in the northeast (n = 4), southeast (n = 6), south (n = 1),

[1] The JRC-DMS, which also accredits sonography educational programs, was considering reorganization at the time of the pilot study. During this same period, ARRT was revising its policies related to the recognition of accrediting agencies. Given this uncertainty, participation was extended only to regionally accredited programs. Note, however, that as of October 2005 the JRC-DMS (through CAAHEP) was recognized by ARRT as an accreditation mechanism acceptable to ARRT, and graduates of sonography programs accredited by JRC-DMS will be eligible for ARRT certification in sonography.

midwest (n = 4), west (n = 2), and northwest (n = 1). Participating students completed the examination between August 4 and September 13, 2005.

To help ensure that students were genuinely motivated, students and programs directors were informed that the ARRT credential in sonography would be awarded to students whose test scores met or exceeded the passing standard. Students were told that they would receive detailed preliminary results at the test center as soon as they completed the exam. Final scores and pass-fail notices were to be mailed on September 30, 2005 at the conclusion of the pilot study, about one week after the standard-setting (i.e., passing score) study had been completed. The standard-setting study is described later in this report.

Instrumentation

The ARRT examination in Sonography consists of 270 questions (250 scored items plus 20 pilots) covering the topics listed in table 1. Two parallel forms of the examination were assembled according to strict specifications regarding the content and difficulty of each section. Students were randomly assigned to take one of the two forms. Each exam form included a set of common items; that is test questions that also appeared on the other form. The use of common items allows the two forms to be placed onto the same score scale using a statistical method known as common item equating. Students were given up to 4.5 hours (270 minutes) to complete the examination.

Table 1. Overview of Content Specifications

	Content Area	**No. of Items**
A.	Patient Care	20
	1. Legal and Ethical Principles	4-6
	2. Patient Monitoring and Safety	5-7
	3. Interpersonal Communications	2-4
	4. Infection Control	5-7
B.	Physical Principles of Ultrasound	80
	1. Generation of Signal	33-37
	2 Propagation of Ultrasound through Tissue	18-22
	3 Image Production	23-27
C.	Sonographic Procedures	150
	1. Abdomen and Pelvis	52-58
	2. Gynecologic Structures	23-27
	3. Obstetrics	52-58
	4. Other (superficial structures, noncardiac chest	14-16
	Experimental Questions	20
	Total	**270**

Establishing a Score Scale

Passing Score Study. A standard-setting workshop was held September 17, 2005 to establish a passing score on the examination. A criterion-referenced procedure known as the Angoff method was used. Fifteen individuals served on the standard-setting panel. Six participants were members of the Examination Committee, with the remaining nine representing various aspects of the professional community. The workshop was facilitated by a Ph.D. psychometrician on ARRT staff.

The standard-setting study resulted in a passing score of 66% correct (165 of 250 questions), which was subsequently approved by the ARRT Board of Trustees. Appendix A presents a one-page summary of the standard-setting workshop; a more detailed technical report is available from ARRT. The raw passing score of 165 was mapped to a scaled passing score of 75 as described below.

Total Scaled Scores. The ARRT uses scaled scores to report exam results. A total scaled score can range from 1 to 99, and a scaled score of 75 is required to pass. Therefore, a raw passing score of 165 (or 66% correct) was mapped to a scaled score of 75 using a linear conversion equation. The conversion equation requires two scaling coefficients: the slope (a) and the intercept (b). The values of a and b are calculated from four values: the maximum scaled score (99.49), the scaled cut score (74.50), the maximum raw score (250) and the raw cut score (165).

$$a = (99.49 - 74.50) / (250 - 165)$$
$$a = 0.294$$

$$b = 74.50 - (a * 165)$$
$$b = 74.50 - (0.294 * 165)$$
$$b = 25.99$$

Therefore, the equation for converting raw scores to scaled scores for the first form is: *scaled = (raw * 0.294) + 25.99*. Scaled scores are desirable because they take into account the difficulty of a particular exam compared to earlier versions of the same exam. Raw scores have limited use because they cannot be compared from one version of an exam to the next, or from one year to the next. This lack of comparability exists because one version of an exam might be slightly easier or slightly more difficult than a previous version. As previously noted, two forms of the exam were used in the pilot study. Through common item equating it was found that the second form was approximately 3 points more difficult than the first form. The scaling equation for the second form was found to be: *scaled = (raw * 0.284) + 28 .563*.

Section Scaled Scores. Performance on the major sections of the exam are also reported using scaled scores. Section scores provide general information about strengths and weaknesses in different content categories. Pass-fail decisions are not made for individual sections of the exam. Section scores can range from 1.0 to 9.9, and are reported at one-tenth point intervals (e.g., 8.1, 8.6). Section scores are placed on a narrower scale because they are based on fewer test items and are less reliable than total scores.

For the pilot study, scores are also reported on a percent correct metric. Percent correct scores are reported for the three major sections and 11 subsections. Reporting scores at this level of specificity is unusual for the ARRT. It is being done for the pilot study to help students and program directors make the most of their score reports. Depending on the type of feedback ARRT receives, it may be decided to continue reporting scores for subsections. Subsection scores warrant a cautious interpretation. First, the subsection scores are not equated. Therefore, scores for subsection B.3 on one form, for example, may not be comparable to scores on subsection B.3 on another form. Second, some subsections contain only a few test items, and scores based on a small collection of items are not very reliable.

3. Results

Seventy-one students from 18 programs completed one of two forms of the examination. Form SOP-01 was administered to a random subset of 36 students, while SOP-02 was administered to a random subset of 35 students. Most analyses are based on equated scaled scores. In certain instances as noted results are reported for percent correct scores.

Reliability and Validity

Score Reliability. Total test reliability, as measured by coefficient α (alpha), was 0.97 for form SOP-1 and 0.94 for form SOP-2. For certification purposes it is important to have values of α that exceed 0.85 and desirable to have values that exceed 0.90. Values obtained from the pilot study indicate that total test scores provide an extremely reliable measure of student proficiency. Reliabilities may be interpreted as a correlation. In this case, the expected correlation of scores on one form of the sonography exam with another parallel sonography exam would in the mid-0.90s. A high correlation such as this means that each person would be expected to earn very similar scores on repeated testings.

The reliability of scores can be more precisely summarized by a margin of error. The *standard error of measurement* is the margin of error used in testing; it represents the variability of a person's score that would be expected on repeated testings. Table 2 indicates the average standard errors of measurement (i.e., the error expected for any score, on average), as well as the standard error at the cut score of 75.

Table 2. Reliability Coefficient and Standard Error of Measurement

Form	Coefficient Alpha	Standard Error of Measurement	
		On Average	At Cut Score
SOP-1	0.97	1.75	1.91
SOP-2	0.94	1.81	1.97

Score reliability was also computed for the three major sections of the exam. Coefficient α for Sections A, B, and C were 0.43, 0.91, and 0.92 when averaged over the two forms. The reliabilities for section scores are not critical because pass-fail decisions are made only at the total score level. However, the values are worthy of consideration when interpreting section scores for purposes related to instructional planning or remediation. Although section scores for sections B and C are quite precise, scores for section A are not very reliable and should be interpreted with caution. The low reliability for section A can be attributed to the fact that it consists of only 20 test items.

One additional noteworthy point is the reliability for Section B (physical principles). An average α of 0.91 (0.92 and 0.91 for the two forms) is quite high for a section consisting of 80 questions. Although the ARRT does not make pass-fail decisions on single content areas, reliability coefficients that exceed 0.90 are generally considered sufficient for making pass-fail decisions.

Validity: Section Score Correlations. The correlations among the section scores provide a measure of their distinctness. Correlations theoretically range from –1.00 (perfect inverse relationship) to +1.00 (perfect positive relationship). In practice, section scores on an exam are usually positively correlated because candidates who perform well on one section typically perform well on others, while those who perform poorly on one section also perform poorly on other sections. The magnitude of a correlation is an index of how related two sections are. Correlations in the 0.40s and 0.50s indicate that sections, although related, measure somewhat unique aspects of candidate knowledge. Correlations in the 0.70s and higher suggest that sections are measuring similar aspects of candidate knowledge. More precisely, a high correlation indicates that a person's performance on one section can be accurately predicted from their scores on another section.

Correlations were computed among the three major sections, as well as for subsections C.1, C.2 and C.3. Subsections C.1 through C.3 are of particular interest because low correlations for a subsection would suggest that: (a) it is a unique content area and should not be included with others sections in the total score, and (b) a separate passing score for that section should be considered. Section C.4 (other) was excluded because it is a heterogeneous content domain consisting of relatively few items, and this inhibits a measure from correlating with any other measure.

Table 3 presents the correlation matrix for the relevant sections. A few key points are:

- Section A (patient care) has correlations lower than those for other sections. There are likely two reasons for this: (a) fewer items in this section result in less reliable section scores, which depresses the correlations; (b) The content domain of patient care is less similar to the other content areas than the other areas are to each other; that is, the subject matter covered in patient care is more unique.
- The correlation of 0.86 between section B and section C is very high, offering strong support for covering both physics and sonographic procedures on a single exam.

Table 3. Correlations of Section Scores

Section	A	B	C	C.1	C.2	C.3
A. Patient Care		0.57	0.62	0.63	0.58	0.56
B. Physical Principles			0.86	0.85	0.73	0.80
C. Sonographic Procedures				0.95	0.89	0.93
1. Abdomen and Pelvis					0.77	0.81
2. Gynecologic Structures						0.79
3. Obstetrics						

- The correlations between section C (sonographic procedures) and C.1, C.2, and C.3 are all very high. However, the magnitude of these correlations is due, in part, to an artifact: C is a composite formed by summing over C.1, C.2, and C.3, and the correlations between a sum and its parts must be high.
- The correlations between C.1, C.2, and C.3 presented in the lower right portion of the table are all high. There is no artifact operating to inflate these values. If anything, the smaller number of items in C.2 (gynecologic structures) will serve to slightly suppress the correlations.

The high correlations among sections C.1, C.2, and C.3 clearly indicate that combining clinical specialties (abdomen, pelvis, obstetrics, gynecology) into a single exam is not only justifiable, but also prudent. The data also support combining section scores into a single total score for purposes of making pass-fail decisions. Additional support for a single pass-fail decision comes from the coefficient alphas reported earlier, which indicate that test items as a whole are measuring the same underlying skill (i.e., proficiency in general sonography), and result in very reliable total scores.

Validity: Student Motivation. Given that this was a pilot study, there was some concern that students would not put forth their best effort. The amount of time required for each student to complete the exam was evaluated to help determine if students were genuinely motivated. The idea here is that if students are not motivated then they will complete the exam in a short amount of time and in a haphazard fashion. This would result in lower scores and these lower scores would be expected to contain much random noise, or measurement error.

Students were allowed up to 270 minutes (4.5 hours), which corresponds to one minute per item. Actual test durations for the pilot sample ranged from 71 minutes to 269 minutes, with a mean of 148 minutes. Table 4 divides the testing time into 30 minute intervals and indicates the percentage of students who completed the exam in that amount of time or less.

Table 4. Percentage of Students Completing the Exam within Specified Time Intervals

	Time in Hours						
	1.5	2.0	2.5	3.0	3.5	4.0	4.5
% Completing	8%	32%	58%	76%	90%	96%	100%

It can be seen that 8% of the students completed the exam in less than 1.5 hours, and that 32% had completed it within 2.0 hours. About one-half of the students completed the exam in 2.5 hours or less. These values are roughly comparable to results for other ARRT exams. For example, 10% of Radiography examinees take 1.5 hours or less to complete a 220 item test. After adjusting for the fact that the sonography exam consists of more test items, it was possible to estimate that students in the sonography pilot study required about 20 minutes less than students who take the Radiography exam.

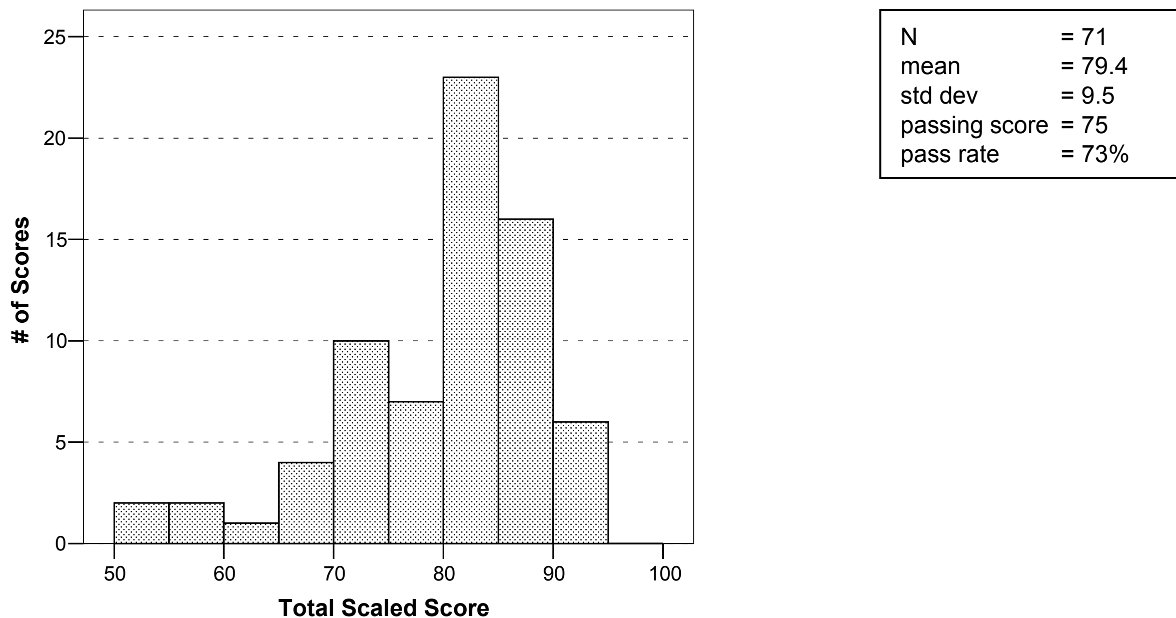
Additional analyses were undertaken to investigate student motivation. It was found that the mean completion time for passers and failers was identical (147.6 min vs. 148.4 min), and there was no

systematic relationship between test scores and test duration (r = 0.02). There were two students whose scores were low enough to warrant some suspicion. Although these two students correctly answered only one-third of the items, they took over three hours to complete the exam. Of all students with failing scores, there were two who completed the exam in less than 90 minutes, perhaps calling into question the motivation of these two students. This concern is offset by the observation that there were also four passers who required less than 90 minutes. The high correlations reported in Table 3 indicate that the section scores contained a high degree of systematic variation. If student's scores were the result guessing or inattention, then one would not expect to see correlations of this magnitude.

Twenty-six students included in the pilot study held previous certification in radiography. Analyses indicated that the radiography scores for these students were slightly higher than average score for all radiography passers (86.5 vs 85.2). In addition, their radiography scores correlated 0.54 with sonography scores, which is comparable to what the relationship seen between radiography and other disciplines (e.g., mammography, MRI). These data suggest that students were capable, as evidenced by their radiography scores, and were generally motivated, as implied by the correlation (r = 0.54). All reasonable checks on the integrity of examination scores indicate that scores from the pilot study are a reliable indicator of student proficiency and lend themselves to valid interpretations regarding student proficiency in general sonography.

Student Results

Total Scores. Total scores on the 250-item exam ranged from 78 (31% correct) to 233 (93% correct). The mean raw score was 181.7 (72.6% correct) with a standard deviation of 32.8 (12.2%). These scores are slightly lower than scores for other ARRT exams, for which mean scores generally range from about 76% to 82% correct. This could suggest either that the sonography exam is more difficult than other exams, or that the sonography students included in the pilot are less well prepared than students in other disciplines.



Raw scores were converted to scaled scores that range from 1 to 99, with 75 defined as a passing point. The figure above presents a bar graph depicting the distribution of scaled scores. The overall pass rate was 73%, with 52 of 71 students passing. Over one-half the scores fall between 80 and 90, with the highest scaled score reaching 94. Of particular interest are the five scores that fell at or below 60. A scaled score of 60 corresponds to about 45% correct, which is a very low score. It seems likely that these

students were not adequately prepared for the exam. Four of these five students were from the same program.

Section Scores. To evaluate how students performed in the various content areas, percent correct and scaled scores were computed for each section of the exam. Table 5 indicates how students performed in the three major content areas. One unanticipated outcome is that the scores for section A (patient care) are lower than expected. Patient care covers those topics of which most health care workers should have basic knowledge. On other ARRT primary exams, average scores on patient care exceed 80% correct.

Table 5. Percent Correct and Scaled Scores on Three Major Sections

	Percent Correct			Scaled Scores		
Section	**low**	**High**	**mean**	**low**	**high**	**mean**
A. Patient Care	35.0	95.0	73.3	5.3	9.6	8.2
B. Physical Principles	28.8	96.3	72.2	4.9	9.7	7.9
C. Sonographic Procedures	30.0	93.3	71.1	5.0	9.5	7.9
Total	30	93.2	72.1	50	94	79.4

The following table presents a more detailed look at performance in specific subsections. Because scaled scores are not computed for subsections, the table presents scores only in the percent correct metric. When interpreting the values in Table 6, it important to keep in mind the small number of questions in some sections.

Table 6. Percent Correct on 11 Subsections

Content Area	# items	low	high	mean
A.1. Legal and Ethical Principles	4-6	0.0	100.0	71.5
A.2. Patient Monitoring and Safety	5-7	14.3	100.0	74.8
A.3. Interpersonal Communications	2-4	33.3	100.0	83.6
A.4. Infection Control	5-7	33.3	100.0	67.6
B.1. Generation of Signal	33-37	20.6	97.2	67.1
B.2 Propagation of Ultrasound through Tissue	18-22	30.8	100.0	78.7
B.3 Image Production	23-27	25.0	95.7	73.2
C.1. Abdomen and Pelvis	52-58	31.0	93.1	72.5
C.2. Gynecologic Structures	23-27	23.1	96.2	69.1
C.3. Obstetrics	52-58	30.2	92.6	72.5
C.4. Other (superficial struct., noncardiac chest)	14-16	30.8	100.0	72.0

There are some notable differences in mean scores. For example, within section A the means range from 83.6 (interpersonal communications) to 67.6 (infection control). Within section B there is about a 10 point difference between the highest and lowest means. It is difficult to ascertain reasons for the differences in scores across different sections. It could be due to students being less prepared in certain content areas, or to some content areas containing more difficult items, or both.

To further explore the reason for low scores in section A (patient care), the pilot sample was divided into two groups: those with prior certification in radiography (n=26) and those without (n=45). It was assumed that because those already certified in radiography would have completed an educational program which covered patient care, and also would have passed the radiography examination which contains a patient care section, then these students should have scores in the 80% correct range (assuming they have not forgotten the material). The mean score for the radiography group on all of section A was 80.2. The scores for those not certified in radiography was 69.3. The mean of 80.2 indicates that the

patient care items really were not overly difficult. These results also suggest that some graduates of sonography programs are not well prepared in patient care.

Student Score Reports. Students received a preliminary score report at the test center upon completing their exam. The printout presented raw (number correct) scores for the 11 subsections and three major sections. Scaled scores and pass-fail outcomes were not provided in the preliminary report. After the pilot study was completed students received a more complete score report that included scaled scores and pass-fail outcomes. Appendix B presents a sample final score report for a passer.

Program Level Results

Total Scores. The 71 students represented 18 educational programs. Summary statistics for each program are provided in Table 7. The table is arranged from highest mean score to lowest. Data are not particularly meaningful for small programs, so Table 7 has been shaded to emphasize those schools with just one or two students. Although extra caution is required when interpreting results for the shaded schools, keep in mind that the *N*s are small for all of the programs. For example, the pass rate for a program with five students would shift by 20% if the outcome for just one student changes from pass to fail or visa versa.

Even when ignoring the smallest programs, it is apparent that there is considerable variability between school means. That is, some programs have very high mean scores and pass rates, while a few have very low mean sores and pass rates. Another interesting outcome is that within a school, the range of scores is remarkably narrow for many of the schools, although there are some exceptions. In particular, score variability for the better performing programs is relatively small.

The finding that several schools had pass rates of 100% could be cause for concern. After ignoring schools with just 1 or 2 students, there remain five schools with pass rates of 100%. Exam results for students from these five schools were further scrutinized by evaluating other data from these schools. As

Table 7. Total Scaled Scores and Pass Rates for Each Program

Program Rank	No. of Students	Min Score	Max Score	Mean Score	Pass Rate
1	1	93	93	93.00	100 %
2	4	85	89	87.25	100 %
3	1	87	87	87.00	100 %
4	2	84	90	87.00	100 %
5	8	77	94	85.75	100 %
6	4	84	86	85.50	100 %
7	5	82	89	84.40	100 %
8	2	82	84	83.00	100 %
9	5	80	86	83.00	100 %
10	7	66	87	78.86	71.4 %
11	2	77	80	78.50	100 %
12	4	67	86	77.75	75 %
13	6	71	83	77.17	66.7 %
14	7	60	91	77.14	57.1 %
15	6	70	83	75.33	33.3 %
16	1	72	72	72.00	0 %
17	2	65	72	68.50	0 %
18	4	50	59	54.50	0 %

it turns out, all five of these schools also offer programs in radiography, and ARRT maintains test performance data for these radiography programs. We know from several years of exam results for numerous educational programs that, in general, if an educational institution is strong in one area of medical imaging (e.g., radiography), then it is likely to be strong in other areas of medical imaging (e.g., nuclear medicine). Similarly, a program that struggles in one specialty is also likely to struggle in others. Although the trend is not perfect, it is seen in other professions, as well.

It was possible to determine from ARRT archival records that three of the five schools with a 100% pass rate in the sonography pilot study also have 100% pass rates in radiography, a result which was confirmed for the past three years. The other two programs have ordinary track records in radiography, but there is no reason to suspect that their students should do poorly in sonography. Assuming that educational programs have strict admission criteria, a thorough curriculum, and qualified faculty, successful outcomes on certification examinations should be expected.

Section Scores. Table 8 presents program means for each major section of the exam. The table also indicates means for the subsections within section C (sonographic procedures), to determine if some programs are particularly strong or weak in just one or two clinical areas. Mean scores are given in the percent correct metric rather than scaled scores because the ARRT computes scaled scores for only for major subsections. As in the previous table, schools are ranked on the basis of their total scores on the exam, and programs with few students have been shaded in gray.

Table 8. Section Scores for Each Program (Percent Correct)

Program Rank	No. of Students	A Patient Care	B Physical Princ	C Sono Procs	C.1 Abd/Pel	C.2 Gyn	C.3 Obstet	C.4 Other
1	1	95.0	92.5	89.3	91.4	88.0	87.0	92.3
2	4	81.3	80.6	84.7	81.0	84.3	90.7	76.9
3	1	80.0	83.8	82.7	84.5	76.0	83.3	84.6
4	2	80.0	84.4	82.3	81.0	84.4	79.4	96.2
5	8	81.3	82.0	80.3	82.1	78.1	80.0	77.9
6	4	81.3	80.3	80.5	78.4	86.4	80.2	78.8
7	5	82.0	81.3	78.1	81.0	77.6	75.6	76.9
8	2	57.5	80.6	78.3	79.3	70.6	78.5	88.5
9	5	74.0	78.5	76.8	78.3	73.5	76.7	76.9
10	7	77.9	70.9	69.8	71.4	71.3	66.4	73.6
11	2	82.5	64.4	74.0	79.3	68.7	72.0	69.2
12	4	70.0	70.9	69.7	65.5	73.1	74.9	59.6
13	6	74.2	69.0	67.8	67.8	58.4	71.2	71.8
14	7	65.0	66.4	70.3	75.6	61.0	68.7	71.4
15	6	64.2	69.4	65.1	64.4	59.4	68.9	64.1
16	1	55.0	72.5	59.3	56.9	60.0	61.1	61.5
17	2	75.0	46.9	60.3	55.2	56.9	68.2	57.7
18	4	48.8	37.5	37.0	36.6	32.8	38.1	42.3

The data in the table indicate that programs tended to have fairly even performance across the different content areas. In particular, each program's means within section C are reasonably consistent. One exception is the program ranked at 13th, which has scores near 70% for sections C1, C.3, and C.4, but only at 58% for C.2 (gynecologic structures). However, for all other programs, it appears as if students

are about equally prepared or unprepared in all areas. These data offer additional support for offering a single, comprehensive examination to program graduates.

Program Director Reports. The ARRT works closely with educational programs to assure that certification outcomes are available and useful to the extent possible. This is done in a manner that protects the confidentiality of each student. Each participating program receives a printout indicating the total scores and section scores for all of his or her students; however, the printout does not present any identifying information, and programs with fewer than three students do not receive the detailed reports. Programs often use such information to evaluate strengths and weaknesses in their curricula or instructional methods. In addition, a graph is provided which compares each program to national norms in 11 subsections and 3 major sections. Appendix C presents a sample report.

4. Summary and Conclusions

A total of 71 students from 18 educational programs participated in the pilot study. Fifty-two students met or exceeded a scaled score of 75 for a pass rate of 73%. The pilot study indicates that an examination consisting of 250 scored questions covering patient care, physics, and various ultrasound procedures produces scores with reliability coefficients that easily exceed commonly accepted psychometric standards. Although the sample size fell just short of expectations, results suggest that students who did participate were motivated to do well. Correlational analyses, as well as an examination of score profiles for each school, support the practice of combining clinical specialties (abdomen, pelvis, obstetrics, gynecology, other) into a single exam with one pass-fail outcome. ARRT certification in sonography will be available to graduates of accredited educational programs beginning in January 2006 at more than 200 test centers nationwide.

Appendices

A. Summary of Standard-Setting Study (1 page)
B. Sample Candidate (Student) Report (1 page)
C. Sample Program Director Reports (2 pages)

Appendix A

Summary of the ARRT Passing Score Study for the Sonography Primary Eligibility Pathway

September 2005

Background

The ARRT's certification program in sonography is being transitioned into a program with a primary pathway. This means that eligibility has been extended to non-RT graduates of accredited educational programs in general sonography. The "go live" date for this transition is January 2006. Sonography certification will also continue to retain its post-primary status.

Because the revised certification program targets a different population, and since the exam required extensive revisions, it is necessary to establish a new cut score, or passing standard, for the exam. A standard setting study was conducted September 17, 2005 at ARRT offices. Fifteen individuals participated in the workshop; six were members of the Examination Committee, two currently serve on other sonography committees (breast and vascular), while the remaining seven represented various facets of the educational community. An ARRT Trustee, also certified in Sonography, served as both workshop participant and Board liaison to the standard-setting panel.

This paper summarizes the standard setting workshop. A more complete report is available from ARRT. Note that a pilot study was also completed. Seventy-one students from 18 accredited educational programs took the revised Examination in Sonography in August or September 2005. Data from the pilot study were used to evaluate the cut score. A separate report on the results of the pilot study will be prepared and shared with participating schools and interested individuals.

Summary of Standard-Setting Workshop

The primary standard-setting procedure used for this study was the Angoff method. This method requires that participants evaluate each individual test item on an examination form and make judgments about the percent of minimally qualified candidates that would be expected to answer the question correctly. After about four hours of orientation and group discussion, the 15 participants provided ratings on the 270 test items (250 scored and 20 unscored) comprising one form of the sonography examination.

Participants also completed a survey consisting of six questions asking for their global impressions of an appropriate cut score and pass rate. Although responses to these six questions can be useful for evaluating the cut score, primary reliance is usually placed on the Angoff ratings.

Standard-Setting Results

Judging from the internal consistency of the ratings, all participants understood the standard-setting process and took their charge seriously. The Angoff-based cut score resulting from the workshop is ***66% correct, or 165 out of 250 questions***. Based on this cut score, the ***pass rate is expected to be about 73%*** for the 71 candidates enrolled in the pilot study. A 73% pass rate is consistent with participants' responses to global survey questions on acceptable pass rates, which suggest that an acceptable pass rate would fall between 66% and 86%.

A standard of 66% correct is consistent with the cut score for other ARRT primary exams, even though the pass rate of 73% for sonography is lower than the pass rate for other ARRT primary exams. However, a relatively low pass rate in sonography might be expected given that the educational infrastructure for the sonography profession is still in its early stages of development and that educational programs vary considerably in scope and quality (e.g., a national curriculum for sonography is still to be developed). The cut score for future forms of the exam will be statistically equated to the exam form used in the standard-setting study.

APPENDIX B

Available on Request from ARRT

Appendix C (page 1 of 2)

ARRT SONOGRAPHY PILOT STUDY SCHOOL RESULTS

School Name Appears Here

Percent Correct in Eleven Content Areas*

Student	A.1	A.2	A.3	A.4	A total	B.1	B.2	B.3	B total	C.1	C.2	C.3	C.4	C total
Student 1	75.0	85.7	66.7	50.0	70.0	80.6	81.0	91.3	83.8	79.3	92.0	92.6	84.6	86.7
Student 2	100.0	100.0	100.0	83.3	95.0	75.0	95.2	82.6	82.5	87.9	88.0	87.0	69.2	86.0
Student 3	75.0	100.0	100.0	50.0	80.0	66.7	90.5	87.0	78.8	77.6	92.0	90.7	84.6	85.3
Student 4	100.0	85.7	100.0	50.0	80.0	76.5	80.8	75.0	77.5	79.3	65.4	92.5	69.2	80.7
Mean	87.5	92.9	91.7	58.3	81.3	74.7	86.9	84.0	80.6	81.0	84.3	90.7	76.9	84.7

Section and Total Scaled Scores

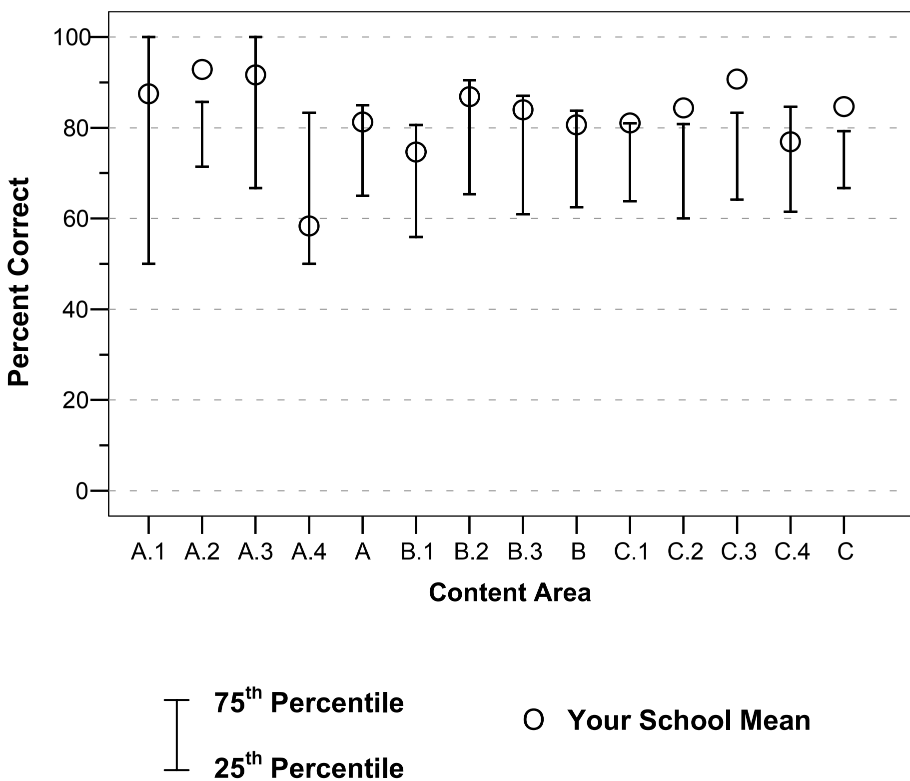
Student	Scale Section A	Scale Section B	Scale Section C	Scaled Total*	Percentile Rank*
Student 1	7.7	8.8	9.0	88	88.0
Student 2	9.6	8.7	8.9	89	90.8
Student 3	8.5	8.4	8.9	87	84.5
Student 4	8.5	8.4	8.6	85	71.8

*Notes: Content areas (A.1, A.2. …) are defined on the following page. A total scaled score of 75 is required to pass. The percentile rank indicates the percentage of students in the pilot sample scoring at or below a given score. Example: If a student scored at the 65th percentile, that means he or she scored better than 64 percent of the pilot sample.

Appendix C (page 2 of 2)
ARRT SONOGRAPHY PILOT STUDY SCHOOL RESULTS

School Name Appears Here

Percent Correct, 3 Major Sections and 11 Subsections



Description of Content Domains (# of questions)

A. Patient Care (20)
1. Legal and Ethical Principles (4-6)
2. Patient Monitoring and Safety (5-7)
3. Interpersonal Communications (2-4)
4. Infection Control (5-7)

B. Physical Principles (80)
1. Generation of Signal (33-37)
2. Propagation of Ultrasound through Tissue (18-22)
3. Image Production (23-27)

C. Sonographic Procedures (150)
1. Abdomen and Pelvis (52-58)
2. Gynecologic Structures (23-27)
3. Obstetrics (52-58)
4. Other Sonographic Exams (14-16)

* Note: Each circle indicates your program's mean in a content domain. The I-bar (vertical line) marks the central portion of the score range based on all 71 students in the pilot study, with the top of the bar corresponding to the 75th percentile and the bottom of the bar corresponding to the 25th percentile.