Annual Exam Report: Technical Appendix - 2023

Introduction

This report summarizes the psychometric characteristics of ARRT's examination scores and is a companion document to the Annual Exam Report published on the exam statistics page of the ARRT website: https://www.arrt.org/pages/arrt-reference-documents/by-document-type/exam-statistics.

For exam security reasons, ARRT regularly publishes new exam forms. However, some of the analyses provided in this report, such as reliability, become more accurate as the number of candidates who see a particular form increases. Therefore, to provide the most accurate information possible, this report will cover exam forms that were predominately offered in 2023 rather than strictly limited to the 2023 calendar year.

The first section of this report contains information about the amount of time that candidates used to complete their examinations. The second section provides descriptive statistics of total exam scores, both raw and scaled, and an explanation of how ARRT converts raw scores to scaled scores. The third section of this report presents descriptive statistics for the exams' section scores, including correlations and reliability estimates. Section four provides more detail about the reliability of the overall exam scores, with a discussion of coefficient α and the standard error of measurement. The final section of the report addresses decision consistency, which quantifies the reproducibility of the certification and registration decisions that ARRT makes based on its examinations.

Information about Exam Durations

Most examination administrators, including ARRT, do not intend for exam administration time to be a major factor for candidates. Practical limitations, however, make it necessary to establish exam time limits. The time limit should allow the exam to begin and end in a reasonable amount of time, while also ensuring that focused and knowledgeable candidates have sufficient time to complete the exam.

Some sources (e.g., Nunnally, 1978) specify that an exam is not speeded when at least 90% of candidates complete the exam within the allotted time. If results show that more than 10% of candidates require the full time, ARRT would consider re-evaluating existing time limits.

Table 1 summarizes the amount of time candidates spent on the exam. These and all other statistics reflect only first-time ARRT exam candidates. None of the statistics include state candidates or people retaking the exam after failing the initial attempt.

Table 1. Time Spent on Examination (in Minutes)

Discipline	Limit	Minimum	Maximum	Mean	Standard Deviation	90th Percentile
Radiography	230	46	230	152	43	216
Nuclear Medicine	230	57	230	160	46	225
Radiation Therapy	230	62	230	180	40	228
Sonography (1)	240	52	237	137	49	214
Sonography (2)	150	28	150	81	30	128
MR Imaging	210	42	210	140	42	201
Mammography	150	22	150	92	28	132
Computed Tomography	195	32	195	116	40	175
Vascular Sonography	225	63	207	129	45	192
Bone Densitometry	105	28	105	69	20	100
Cardiac Interventional	195	49	195	123	38	185
Vascular Interventional	210	53	210	139	39	195
Breast Sonography	225	29	224	134	43	199
Radiologist Assistant (1)	240	58	234	156	47	216
Radiologist Assistant (2)	120	50	117	66	23	94

Note: The Sonography and Radiologist Assistant exams are split into two sessions with separate time limits.

Descriptive Statistics for Total Examination Scores

Table 2 contains descriptive statistics for the raw scores (number correct), which are the basis for numerous other calculations in this report. The unscored “pilot” items used for testing are not included.

Table 2. Raw Score by Discipline

Discipline	Scored	Minimum	Maximum	Mean	Standard Deviation
Radiography	200	53	199	154.7	20.8
Nuclear Medicine	200	62	194	147.9	23.5
Radiation Therapy	200	66	194	155.3	18.5
Sonography	360	110	343	256.5	43.2
MR Imaging	200	52	197	147.2	24.6
Mammography	115	34	114	90.3	11.6
Computed Tomography	165	58	162	119.7	17.8
Vascular Sonography	175	69	151	114.5	23.9
Bone Densitometry	75	30	72	56.4	8.2
Cardiac Interventional	145	56	130	94.3	15.8
Vascular Interventional	160	53	152	111.8	16.1
Breast Sonography	185	41	176	142.4	19.4
Radiologist Assistant*	266	138	239	185.5	26.7

*The Radiologist Assistant Exam consists of two sections with slightly different raw score interpretations. The primary section includes 200 scored items and accounts for 200 points (75% of the total). The case study section includes at least 30 scored items, and the remaining 66 points

(25% of the total) are divided equally amongst them. The number of scored items in the case study section varies based on the cases selected for that administration. The values presented here are based on the 266 total points available to facilitate comparison across different exam forms.

ARRT uses scaled scores to report exam results. Total scaled scores range from 1 to 99, and a candidate must achieve a total scaled score of 75 to pass an examination. Table 3 contains descriptive statistics for the total scaled scores. The main advantage of scaled scores is that they allow readers to compare scores across forms and years.

Each exam consists of items that were used on previous exams. ARRT uses the Rasch model to track the difficulty levels of individual exam items and, in aggregate, whole exam forms. Each item has a Rasch difficulty statistic indicating the probability of a candidate answering correctly.

ARRT determines the difficulty of an exam form by calculating the sum of the probabilities of correct answers at the cutpoint. Comparisons with the difficulties of previous forms determine the relative difficulty level of the new form. If the new form is easier, the cut score for the new form will be greater by an appropriate number of questions. If the new form is more difficult, then the cut score will be lower by some appropriate number of questions.

After determining the raw passing score, ARRT calculates equations to convert the raw scores to scaled scores such that the scaled scores range from 1 to 99 with a passing score of 75. As a hypothetical example, assume that the raw passing score is 130 out of 200. The conversion equation requires two scaling coefficients: the slope (a) and the intercept (b). The calculations of a and b involve four values: the maximum scaled score (99.49), the scaled cut score (74.50), the maximum raw score (200), and the raw cut score (130).

$$a = (99.49 - 74.50) / (200 - 130) = 0.357$$

$$b = 74.50 - (a \times 130) = 74.50 - (0.357 \times 130) = 28.09$$

For this hypothetical form, the scaling coefficients would be $a = 0.357$ and $b = 28.09$. ARRT would use these scaling coefficients to convert the raw scores to scaled scores. If a candidate achieved a raw score of 131 (one point above passing), then the scaled score would be

$$\text{scaled score} = (\text{raw score} \times 0.357) + 28.09 = (131 \times 0.357) + 28.09 = 74.857,$$

which rounds up to 75, a passing scaled score. For this example, raw scores of 130 and 131 round up to a passing scaled score of 75. Raw scores of 128 and 129, however, round to a scaled score of 74, which is a failing score.

Table 3. Scaled Score by Discipline

Discipline	Minimum	Maximum	Mean	Standard Deviation
Radiography	43	99	82.0	8.0
Nuclear Medicine	52	97	81.7	8.1
Radiation Therapy	47	97	81.8	7.3
Sonography	46	96	77.1	9.3
MR Imaging	44	98	79.6	9.2
Mammography	47	99	83.6	7.5
Computed Tomography	51	98	79.1	8.0
Vascular Sonography	55	89	73.2	10.3
Bone Densitometry	50	96	78.9	9.1
Cardiac Interventional	55	92	75.6	8.0
Vascular Interventional	46	95	74.7	8.8
Breast Sonography	41	96	82.4	8.2
Radiologist Assistant	64	92	77.1	7.3

Descriptive Statistics for Section Scores

In addition to the total scaled score, ARRT reports individual section scores that correspond to content areas as outlined in the content specifications of each exam. These section scores are intended to provide general information to candidates regarding their strengths and weaknesses in particular content categories. For Sonography only, candidates must pass both the Abdomen and OB/GYN sections with a section scaled score of 7.5 in addition to passing the exam as a whole. ARRT reports section scores on a scale from 0.1 to 9.9 in one-tenth point intervals.

Section scores are useful to the extent that: (a) the scores are reliable and (b) the sections measure knowledge and skills that are independent of each other. For these reasons, Tables 4 through 16 contain additional descriptive statistics about ARRT's section scores. These include the correlations among the section scores as well as section score means and standard deviations. In addition, the tables contain a reliability estimate (Cronbach's α) for each section. Sections with more items generally have more reliable scores in the same way that longer examinations generally have more reliable scores. Reliability is discussed in more detail later in this report.

The correlations among the section scores provide a measure of their distinctness. In theory, correlations can range from –1.00 (perfect inverse linear relationship) to +1.00 (perfect positive linear relationship). Section scores on an exam are usually positively correlated because candidates who perform well on one section typically perform well on others. For Tables 4 through 16, the upper panel displays correlations among section scores. When interpreting the correlations in Tables 4 through 16, it is important to consider the reliability of each section score. Sections with low reliability will naturally show lower correlations with other subscales. A low reliability coefficient for a section also indicates that a candidate's score for that section is only an approximation of the candidate's true level of knowledge. For this reason, ARRT cautions students and program directors not to over-interpret small score differences among section scores. The limited reliability of section scores is the primary reason that ARRT bases its pass/fail decisions on total scores. Total scores are sufficiently reliable to make pass/fail decisions; section scores may not have sufficient reliability to make those decisions. A notable exception to this is Sonography. ARRT does base pass/fail decisions on the Abdomen and OB/GYN sections of that exam, and the reliability of those section scores is quite high.

Table 4. Radiography (RAD) Section Correlation and Statistics

Section	PC1	S1	S2	IP1	IP2	P1	P2	P3
PC1	*							
S1	0.52	*						
S2	0.53	0.60	*					
IP1	0.53	0.63	0.66	*				
IP2	0.53	0.62	0.61	0.65	*			
P1	0.44	0.54	0.55	0.56	0.53	*		
P2	0.54	0.55	0.56	0.57	0.54	0.56	*	
P3	0.51	0.56	0.58	0.60	0.55	0.60	0.58	*
Statistic								
Mean Scaled	8.2	8.2	8.2	8.3	8.3	8.1	8.0	8.2
SD Scaled	0.9	1.1	1.0	1.1	1.0	1.2	1.1	1.0
Reliability	0.62	0.64	0.68	0.69	0.61	0.58	0.60	0.64

Section	Name	Items
	Patient Care	
PC1	*Patient Interactions and Management*	33
	Safety	
S1	*Radiation Physics and Radiobiology*	21
S2	*Radiation Protection*	29
	Image Production	
IP1	*Image Acquisition and Technical Evaluation*	26
IP2	*Equipment Operation and Quality Assurance*	25
	Procedures	
P1	*Head, Spine, and Pelvis Procedures*	18
P2	*Thorax and Abdomen Procedures*	20
P3	*Extremity Procedures*	28

Table 5. Nuclear Medicine Technology (NMT) Section Correlation and Statistics

Section	PC1	S1	IP1	P1	P2	P3	P4	P5
PC1	*							
S1	0.47	*						
S2	0.58	0.69	*					
IP1	0.47	0.67	0.69	*				
IP2	0.54	0.60	0.64	0.64	*			
P1	0.50	0.62	0.66	0.65	0.66	*		
P2	0.45	0.53	0.61	0.60	0.62	0.62	*	
P3	0.54	0.65	0.68	0.68	0.67	0.68	0.64	*
Statistic								
Mean Scaled	8.5	8.1	7.9	8.1	8.2	8.3	8.3	8.1
SD Scaled	0.8	1.0	1.0	1.0	1.0	1.0	1.0	1.1
Reliability	0.67	0.72	0.78	0.75	0.75	0.73	0.67	0.68

Section	Name	Items
	Patient Care	
PC1	*Patient Interactions and Management*	24
	Safety	
S1	*Radiation Physics, Radiobiology, and Regulations*	25
	Image Production	
IP1	*Instrumentation*	33
	Procedures	
P1	*Radionuclides and Radiopharmaceuticals*	28
P2	*Cardiac Procedures*	25
P3	*Endocrine and Oncology Procedures*	25
P4	*Gastrointestinal and Genitourinary Procedures*	18
P5	*Other Imaging Procedures*	22

Table 6. Radiation Therapy (THR) Section Correlation and Statistics

Section	PC1	PC2	S1	S2	P1	P2	P3	P4
PC1	*							
PC2	0.47	*						
S1	0.52	0.52	*					
S2	0.54	0.46	0.59	*				
P1	0.45	0.50	0.55	0.49	*			
P2	0.42	0.48	0.51	0.53	0.51	*		
P3	0.48	0.47	0.56	0.53	0.49	0.47	*	
P4	0.52	0.52	0.60	0.62	0.56	0.56	0.58	*
Statistic								
Mean Scaled	8.5	8.4	8.0	8.1	8.1	8.4	7.9	8.1
SD Scaled	0.8	1.0	1.2	0.9	0.9	1.0	1.0	0.9
Reliability	0.51	0.46	0.62	0.62	0.58	0.45	0.60	0.68

Section	Name	Items
	Patient Care	
PC1	*Patient Interactions and Management*	29
PC2	*Patient and Medical Record Management*	17
	Safety	
S1	*Radiation Physics, Equipment Operation, and Quality Assurance*	21
S2	*Radiation Protection*	30
	Procedures	
P1	*Treatment Sites and Tumors*	26
P2	*Treatment Volume Localization*	18
P3	*Prescription and Dose Calculation*	24
P4	*Treatments*	35

Table 7. Sonography (SON) Section Correlation and Statistics

Section	PC1	IP1	IP2	IP3	P1	P2	P3
PC1	*						
IP1	0.45	*					
IP2	0.46	0.80	*				
IP3	0.46	0.69	0.65	*			
P1	0.54	0.64	0.61	0.61	*		
P2	0.52	0.62	0.62	0.61	0.84	*	
P3	0.57	0.54	0.54	0.55	0.76	0.75	*
Statistic							
Mean Scaled	7.4	7.7	7.6	7.8	7.8	7.8	7.6
SD Scaled	1.1	1.1	1.1	1.2	1.1	1.1	1.1
Reliability	0.69	0.84	0.82	0.53	0.89	0.93	0.60

Section	Name	Items
	Patient Care	
PC1	*Patient Interactions and Management*	29
	Image Production	
IP1	*Basic Principles of Ultrasound*	50
IP2	*Image Formation*	44
IP3	*Evaluation and Selection of Representative Images*	21
	Procedures	
P1	*Abdomen*	75
P2	*Obstetrics and Gynecology*	109
P3	*Superficial Structures and Other Sonographic Procedures*	32

Note: The Obstetrics and Gynecology score is a composite of the First Trimester Obstetrics, Second/Third Trimester and High Risk Obstetrics, and Gynecology sections.

Table 8. MR Imaging (MRI) Section Correlation and Statistics

Section	PC1	S1	IP1	IP2	IP3	P1	P2	P3
PC1	*							
S1	0.46	*						
IP1	0.46	0.64	*					
IP2	0.40	0.61	0.80	*				
IP3	0.45	0.60	0.77	0.78	*			
P1	0.47	0.57	0.65	0.66	0.64	*		
P2	0.44	0.46	0.51	0.48	0.51	0.58	*	
P3	0.42	0.50	0.55	0.55	0.53	0.61	0.52	*
Statistic								
Mean Scaled	8.1	8.2	8.0	8.0	7.5	8.1	7.7	8.0
SD Scaled	1.0	1.0	1.1	1.2	1.2	1.1	1.2	1.1
Reliability	0.45	0.62	0.82	0.84	0.76	0.72	0.56	0.48

Section	Name	Items
	Patient Care	
PC1	*Patient Interactions and Management*	18
	Safety	
S1	*MRI Screening and Safety*	20
	Image Production	
IP1	*Physical Principles of Image Formation*	39
IP2	*Sequence Parameters and Options*	36
IP3	*Data Acquisition, Processing, and Storage*	30
	Procedures	
P1	*Neurological*	25
P2	*Body*	15
P3	*Musculoskeletal*	17

Table 9. Mammography (MAM) Section Correlation and Statistics

Section	PC1	IP1	P1	P2
PC1	*			
IP1	0.52	*		
P1	0.54	0.62	*	
P2	0.56	0.61	0.63	*
Statistic				
Mean Scaled	8.4	8.3	8.5	8.3
SD Scaled	1.1	0.9	0.9	0.8
Reliability	0.48	0.67	0.64	0.68

Abbreviation	Section Name	Items
	Patient Care	
PC1	*Patient Interactions and Management*	14
	Image Production	
IP1	*Image Acquisition and Quality Assurance*	33
	Procedures	
P1	*Anatomy, Physiology, and Pathology*	26
P2	*Mammographic Positioning, Special Needs, and Imaging Procedures*	42

Table 10. Computed Tomography (CT) Section Correlation and Statistics

Section	PC1	S1	IP1	IP2	P1	P2	P3
PC1	*						
S1	0.44	*					
IP1	0.53	0.57	*				
IP2	0.52	0.58	0.64	*			
P1	0.48	0.49	0.53	0.53	*		
P2	0.53	0.50	0.53	0.56	0.56	*	
P3	0.52	0.48	0.50	0.55	0.55	0.62	*
Statistic							
Mean Scaled	8.1	8.2	7.8	7.9	7.9	7.9	7.7
SD Scaled	0.9	1.0	1.0	1.1	1.0	1.0	1.1
Reliability	0.50	0.60	0.65	0.64	0.61	0.59	0.65

Section	Name	Items
	Patient Care	
PC1	*Patient Interactions and Management*	22
	Safety	
S1	*Radiation Safety and Dose*	22
	Image Production	
IP1	*Image Formation*	28
IP2	*Image Evaluation and Archiving*	22
	Procedures	
P1	*Head, Spine, and Musculoskeletal*	25
P2	*Neck and Chest*	21
P3	*Abdomen and Pelvis*	25

Table 11. Vascular Sonography (VS) Section Score Correlation Matrix and Statistics

Section	PC1	IP1	IP2	IP3	P1	P2	P3	P4
PC1	*							
IP1	0.17	*						
IP2	0.13	0.58	*					
IP3	0.12	0.67	0.65	*				
P1	0.28	0.32	0.26	0.55	*			
P2	0.04	0.73	0.47	0.76	0.35	*		
P3	0.32	0.70	0.54	0.82	0.52	0.70	*	
P4	0.06	0.49	0.31	0.71	0.45	0.81	0.70	*
Statistic								
Mean Scaled	7.2	7.3	7.4	7.3	7.2	7.9	7.1	7.3
SD Scaled	1.3	1.1	1.4	1.5	1.2	1.4	1.6	1.4
Reliability	0.68	0.74	0.67	0.85	0.65	0.79	0.74	0.71

Abbreviation	Section Name	Items
	Patient Care	
PC1	*Patient Interactions and Management*	20
	Image Production	
IP1	*Basic Principles of Ultrasound*	26
IP2	*Image Formation*	14
IP3	*Evaluation and Selection of Representative Images*	35
	Procedures	
P1	*Abdominal/Pelvic Vasculature*	23
P2	*Arterial Peripheral Vasculature*	21
P3	*Venous Peripheral Vasculature*	18
P4	*Extracranial Cerebral Vasculature and Other Sonographic Procedures*	18

Table 12. Bone Densitometry (BD) Section Score Correlation Matrix and Statistics

Section	PC1	IP1	P1
PC1	*		
IP1	0.47	*	
P1	0.57	0.57	*
Statistic			
Mean Scaled	8.0	7.7	7.9
SD Scaled	1.1	1.2	1.0
Reliability	0.62	0.64	0.72

Section	Name	Items
	Patient Care	
PC1	*Patient Bone Health, Care, and Radiation Principles*	17
	Image Production	
IP1	*Equipment Operation and Quality Control*	20
	Procedures	
P1	*DXA Scanning*	38

Table 13. Cardiac Interventional (CI) Section Score Correlation Matrix and Statistics

Section	PC1	IP1	P1	P2
PC1	*			
IP1	0.62	*		
P1	0.64	0.53	*	
P2	0.68	0.61	0.74	*
Statistic				
Mean Scaled	7.5	7.5	7.4	7.6
SD Scaled	0.9	0.9	1.0	1.0
Reliability	0.62	0.49	0.71	0.77

Section	Name	Items
	Patient Care	
PC1	*Patient Interactions and Management*	35
	Image Production	
IP1	*Image Acquisition and Equipment*	30
	Procedures	
P1	*Diagnostic and Conduction System Procedures*	36
P2	*Hemodynamics, Calculations, and Percutaneous Intervention*	44

Table 14. Vascular Interventional (VI) Section Score Correlation Matrix and Statistics

Section	PC1	IP1	P1	P2	P3
PC1	*				
IP1	0.43	*			
P1	0.58	0.56	*		
P2	0.55	0.54	0.74	*	
P3	0.48	0.54	0.58	0.60	*
Statistic					
Mean Scaled	7.5	7.4	7.3	7.5	7.7
SD Scaled	1.0	1.0	1.2	1.0	1.1
Reliability	0.55	0.59	0.78	0.74	0.66

Section	Name	Items
	Patient Care	
PC1	*Patient Interactions and Management*	30
	Image Production	
IP1	*Image Acquisition and Equipment*	25
	Procedures	
P1	*Vascular Diagnostic Procedures*	50
P2	*Vascular Interventional Procedures*	35
P3	*Nonvascular Procedures*	20

Table 15. Breast Sonography (BS) Section Score Correlation Matrix and Statistics

Section	PC1	IP1	IP2	IP3	P1	P2	P3
PC1	*						
IP1	0.57	*					
IP2	0.58	0.70	*				
IP3	0.61	0.71	0.71	*			
P1	0.50	0.60	0.53	0.56	*		
P2	0.68	0.63	0.62	0.63	0.55	*	
P3	0.50	0.40	0.48	0.47	0.37	0.55	*
Statistic							
Mean Scaled	8.4	7.7	8.3	8.3	8.4	8.4	8.5
SD Scaled	0.8	1.1	0.9	1.0	1.1	0.8	1.0
Reliability	0.79	0.92	0.95	0.95	0.82	0.93	0.97

Abbreviation	Section Name	Items
	Patient Care	
PC1	*Patient Interactions and Management*	18
	Image Production	
IP1	*Basic Principles of Ultrasound*	37
IP2	*Image Formation*	32
IP3	*Evaluation and Selection of Representative Images*	33
	Procedures	
P1	*Anatomy and Physiology*	15
P2	*Pathology*	35
P3	*Breast Interventions*	15

Table 16. Radiologist Assistant (RA) Section Score Correlation Matrix and Statistics

Section	PC1	PC2	S1	P1	P2	P3	P4	Case
PC1	*							
PC2	0.72	*						
S1	0.57	0.49	*					
P1	0.66	0.62	0.54	*				
P2	0.69	0.64	0.56	0.74	*			
P3	0.60	0.61	0.62	0.56	0.64	*		
P4	0.68	0.63	0.58	0.57	0.70	0.66	*	
Case	0.66	0.65	0.59	0.68	0.81	0.77	0.63	*
Statistic								
Mean Scaled	8.2	7.3	7.6	8.0	7.5	7.0	7.7	7.7
SD Scaled	0.6	1.4	0.7	0.9	0.9	1.2	1.0	0.7
Reliability	0.41	0.71	0.23	0.70	0.32	0.62	0.45	0.37

Section	Name	Items
	Patient Care	
PC1	*Patient Management*	38
PC2	*Pharmacology*	18
	Safety	
S1	*Patient Safety, Radiation Protection, and Equipment Operation*	28
	Procedures	
P1	*Abdominal Section*	41
P2	*Thoracic Section*	25
P3	*Musculoskeletal and Endocrine Sections*	25
P4	*Neurological, Vascular, and Lymphatic Sections*	25
Case	Case Study Selection	30+

Reliability of Exam Scores

Reliability refers to the repeatability and consistency of exam scores. A candidate who takes one form of an exam on one occasion and a second parallel form on another occasion should earn similar scores if the exam scores are reliable and the candidate has not changed in the time between the exam administrations (i.e., learned new material). Major differences should occur only if there is true change in the candidate's knowledge or if the exam scores are unreliable.

Reliability also describes how well candidates' observed scores on an exam approximate their "true" scores. A candidate's true score may be defined as the mean of their observed scores from a large number of examinations. The true score is theoretical and not observable in practice.

Reliability coefficients are estimates of the reliability of exam scores. Reliability coefficients typically range from zero to one, with values near one indicating high consistency and those near zero indicating little or no consistency. In this report, Cronbach's coefficient α is the reliability estimate of choice. Cronbach's α, which requires only one exam administration, is an estimate of the reliability of a group's exam scores. Although it is never possible to determine the exact amount of error in one specific candidate's score, the standard error of measurement (SEM) describes the expected variation of each candidate's observed score around that candidate's true score.

Coefficient Alpha

The equation for Cronbach's coefficient *α* is

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{I}\hat{\sigma}_i^2}{\hat{\sigma}_X^2}\right), \tag{1}$$

where *k* is the number of items,
I is the total number of items,
X is a set of exam scores,
$\hat{\sigma}_i^2$ is the variance on an individual item *i*, and
$\hat{\sigma}_X^2$ is the total exam variance.

Standard Error of Measurement

The standard error of measurement (SEM) is a type of standard deviation. SEM is the standard deviation of a hypothetical set of repeated measurements for a single individual. A common equation calculates the SEM using the reliability estimate, r_{XX} (α from Equation 1), and the standard deviation of exam scores, S_X, with the equation

$$\text{SEM} = S_X\sqrt{1 - r_{XX}} \tag{2}$$

The above equation for SEM represents the mean SEM across all exam scores. SEM is not consistent, however, across the full range of scores, especially at the extremes. The SEM calculated at the cut score and the mean score will give a more accurate picture of the standard error. The equation for SEM at a particular score is

$$\mathrm{SEM}_{\hat{X}} = \sqrt{\left(\frac{\hat{X}(k-\hat{X})}{k-1}\right)\left(\frac{1-r_{XX}}{1-r_{21}}\right)}, \quad (3)$$

where $\hat{X}$ is a score value of interest,
k is the number of items,
r_{XX} is the reliability of scores using Cronbach's *α*, and
r_{21} is the reliability of scores using Kuder-Richardson Equation 21 (Lord, 1955; Keats, 1957).

Table 17 provides the standard error of measurement for the mean score and the cut score in both raw and scaled score units using Equation 3

Table 17. Mean Indices of Internal Consistency and Standard Error of Measurement

Discipline	Minimum α	SEM at Mean Score		SEM at Cut Score	
		Raw	Scaled	Raw	Scaled
Radiography	0.93	5.70	2.20	6.37	2.46
Nuclear Medicine	0.95	6.43	2.20	7.08	2.41
Radiation Therapy	0.91	5.67	2.24	6.33	2.50
Sonography	0.97	8.13	1.76	8.39	1.81
MR Imaging	0.95	5.77	2.16	6.18	2.32
Mammography	0.87	4.24	2.73	4.89	3.15
Computed Tomography	0.91	5.37	2.42	5.68	2.56
Vascular Sonography	0.95	5.75	2.49	5.68	2.46
Bone Densitometry	0.84	3.55	3.93	3.74	4.15
Cardiac Interventional	0.89	5.27	2.64	5.25	2.63
Vascular Interventional	0.91	5.39	2.76	5.38	2.76
Breast Sonography	0.99	5.26	2.13	5.86	2.37
Radiologist Assistant	0.91	7.77	2.14	8.14	2.25

Decision Consistency

ARRT administers examinations with criterion-referenced cut score standards as the basis of decisions to grant certification and registration. Agreement indices quantify the consistency or reproducibility of those dichotomous (two option) decisions. Decision consistency in this case describes how consistently the examinations classify individuals into certified and registered and not certified and registered groups. When organizations base a pass/fail decision on a single exam score, there will be a small number of candidates who passed but should have failed (false positives) and a small number of candidates who failed but should have passed (false negatives). The threshold loss agreement indices used in this report focus on the consistency of classifications, treating all potential misclassification errors as equally serious.

The threshold loss indices assume a dichotomous, qualitative classification of candidates as certified and registered or not certified and registered based on a cut score. The methods were originally developed using two or more exam administrations for every candidate. Because

multiple examinations are not practical, researchers developed alternative methods to estimate the indices with a single exam administration. This report uses a method developed by Subkoviak (1976) to estimate two threshold loss indices, p_0 and kappa. The estimation procedure assumes that a candidate's observed scores are independently and binomially distributed according to the number of exam items and the candidate's proportion-correct true score.

p_0 index

The p_0 index measures the overall consistency of pass/fail classifications. It is the proportion of individuals expected to be consistently classified as certified and registered and not certified and registered based on Subkoviak's (1976) method. The index is sensitive to the cut score, exam length, and score variability. For example, p_0 values will be smaller for cut scores near the mean of scores, because there are more people located near the mean than at the extremes if scores are normally distributed. The first column in Table 18 contains the p_0 values for each of the exams that this report covers. Classification decisions based on these exams are consistent between 83% and 93% of the time. This is a high level of decision consistency.

Table 18. Threshold Loss Indices

Discipline	p_0	p_c	kappa
Radiography	0.93	0.74	0.73
Nuclear Medicine	0.93	0.71	0.76
Radiation Therapy	0.92	0.75	0.70
Sonography*	0.93	0.54	0.85
Magnetic Resonance Imaging	0.91	0.62	0.77
Mammography	0.92	0.79	0.60
Computed Tomography	0.89	0.61	0.72
Vascular Sonography	0.90	0.50	0.79
Bone Densitometry	0.84	0.58	0.63
Cardiac Interventional	0.83	0.50	0.66
Vascular Interventional	0.85	0.50	0.70
Breast Sonography	0.92	0.73	0.71
Radiologist Assistant	0.91	0.56	0.80

* The p_0 statistic for SON makes a statistical adjustment to Subkoviak's (1976) method that accounts for the necessity to pass the overall exam, the Abdomen section, and the OB/GYN section.

Kappa

While high classification consistencies are good, it is possible that some or many of the correct classifications of certified and registered or not certified and registered were due to chance. For example, a person can correctly guess heads or tails at the flip of a coin a certain percentage of the time. These correct guesses are due purely to chance. Kappa is a statistical index that shows the proportion of individuals consistently classified beyond that expected by chance. The equation for kappa is

$$k = \frac{p_0 - p_c}{1 - p_c}, \tag{4}$$

where p_0 is the overall consistency of certified and registered/not certified and registered classifications and p_c is the proportion of consistent classifications that would be expected by chance.

The calculation for p_c is simply

$$p_c = (P_{Pass})^2 + (1 - P_{Pass})^2, \quad (5)$$

where P_{pass} is the proportion of people who pass the exam (Croker & Algina, 1986). Table 18 contains the kappa statistics for ARRT's exams. The kappa coefficient indicates that ARRT's exams consistently classify between 60% and 85% of the candidates above and beyond those already correctly classified by chance.

With regard to psychometric properties, ARRT's examinations are comparable to other well-developed examinations. The threshold loss indices indicate that most candidates are consistently classified as either certified and registered or not certified and registered. Maintaining a high-quality examination program is a vital part of ARRT's mission of promoting high standards of patient care by recognizing qualified individuals in medical imaging, interventional procedures, and radiation therapy. The results from this technical report show that ARRT indeed continues to develop quality examinations.

Updates

As of 2024, the Technical Appendix now includes all ARRT certification exams.

Updated content specifications for the Radiologist Assistant exam went into effect January 2023. Notably, these content specifications replaced the case study essay responses with objectively scored items. You can read more in the Updated R.R.A. Documents news release or the exam content specifications page at ARRT.org. With the discontinuation of essay response items, Radiologist Assistant exam statistics will be presented in this annual report and the technical appendix rather than as a separate publication.

Updated content specifications went into effect July 2023 for both Cardiac Interventional and Vascular Interventional Radiography. You can read more in the Updated Cardiac Interventional Radiography Documents and Updated Vascular Interventional Radiography Documents news releases at ARRT.org. Due to limited activity in 2023, performance data related to the new content specifications will be available in the next report.

References

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Belmont, CA: Wadsworth.

Keats, J.A. (1957). Estimation of error variances of test scores. *Psychometrika, 2*, 29-41.

Lord, F.M. (1955). Estimating test reliability. *Educational and Psychological Measurement, 15*, 325-336.

Nunnally, J.C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Subkoviak, M.J. (1976). Estimating reliability from a single administration of a mastery test. *Journal of Educational Measurement, 13*, 265-276.