# Table of Contents

Executive Summary ............................................................................................................. 2

Test Specifications .................................................................................................................. 2

Purpose .................................................................................................................................. 2

Intended Population .............................................................................................................. 3

Blueprint ................................................................................................................................. 3

Examination Overview ........................................................................................................ 3

Subject Matter Experts ........................................................................................................ 3

Item Development ................................................................................................................... 4

Item Formats ............................................................................................................................ 4

Administration ......................................................................................................................... 5

Use of Scores ........................................................................................................................... 5

Construct Irrelevant Variance ............................................................................................... 6

Fairness Review ....................................................................................................................... 6

Scoring Process ....................................................................................................................... 6

Item Analysis and Key Validation ...................................................................................... 6

Classical and IRT Scoring ...................................................................................................... 6

Results .................................................................................................................................... 6

ANOVA .................................................................................................................................. 7

Results .................................................................................................................................... 7

Item Resident Distribution .................................................................................................... 9

Reliability ............................................................................................................................... 11

ABOS/AAOS Linking Study ................................................................................................... 11

Linking Study Design ........................................................................................................... 12

Results .................................................................................................................................... 12

Recommendations .................................................................................................................. 13

Summary ................................................................................................................................. 14

Reference List ......................................................................................................................... 15

Appendix A 2021 Education Assessments and Examinations Committee and Contributors ................................................................................................................................. 16

Appendix B Letter from the Assessments and Examinations Committee Chair ........... 21
Executive Summary

The 2021 Orthopaedic In-Training Examination (OITE) was administered to 4,987 residents across 231 national and international orthopaedic residency programs from November 5, 2021 through November 14, 2021. The OITE is a 275-item, multiple choice, computer-based examination that covers 10 content domains representative of the established principles and conventional procedures and treatment modalities in orthopaedic surgery. Historically, the OITE was administered at designated locations by each residency program in a proctored environment. In 2021, the AAOS hosted two proctored administration models: in-person group testing and remote testing.

Scoring and reporting are conducted by the American Academy of Orthopaedic Surgeons and results are made available to the residency program directors for dissemination to the residents. Score reports include overall program performance, program year (PGY) performance and examinee performance. In 2021, the AAOS adopted a new standard for performance comparisons at all levels (program, PGY, and individual) to better align with the direction of education in US orthopaedic residency programs. To maintain equity in the comparison of performance outcomes and percentiles, the Accreditation Council for Graduate Medical Education (ACGME)-accredited orthopaedic residency programs were used as the reference group. All score comparisons and percentiles point to the ACGME-accredited program results and all programs (domestic and international) are compared to the same reference group.

In 2021, the AAOS and the American Board of Orthopaedic Surgery (ABOS) collaborated on the development of a collection of examination items that were included on both the 2021 ABOS Part I Certifying Examination and the 2021 AAOS OITE. The purpose of including a set of common items on both examinations was to identify the score on the 2021 OITE that approximately corresponds to the minimum passing performance level on the ABOS Part I Certifying Examination. The score estimate is based on a relatively small sample of shared items and is not intended to be predictive of future performance on the ABOS Part I Examination but serve as a rough benchmark to help guide education and examination preparation.

Results from the 2021 administration show a progression of knowledge as the residents advance in their training. Composite mean scores increase significantly across the PGY program (see ANOVA).

Test Specifications

Purpose

The purpose of the Orthopaedic In-Training Examination (OITE) is to provide evidence of the orthopaedic surgeon’s scope of knowledge throughout their training in support of educational advancement. Since 1963, the American Academy of Orthopaedic Surgeons (AAOS) has developed the Orthopaedic In-Training Examination (OITE) to
assess resident knowledge in ten primary content domains as defined by the OITE blueprint.

**Intended Population**

The OITE is made available to all United States Accreditation Council for Graduate Medical Education (ACGME) and American Osteopathic Association (AOA) accredited orthopaedic residency programs as well as Canadian and International (outside North America) programs by request.

The examination is utilized throughout the 5 post-graduate years (PGY) representative of the ACGME training criteria. However, among other groups (International), the number of PGYs may vary.

**Blueprint**

The 2021 OITE content consisted of 10 primary subject matter domains as outlined in the blueprint (Table 1). The content distribution aligned with historical examinations and is reviewed annually by the Education Assessments and Examinations Committee. The AAOS has adopted a new blueprint, developed by the American Board of Orthopaedic Surgery, that will be in place for 2021 test development and examination administration.

**Table 1 2021 OITE Content Domains and Distribution**

<table>
<thead>
<tr>
<th>Domain</th>
<th>Proportion of Scored Exam</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Science</td>
<td>11%</td>
</tr>
<tr>
<td>Foot &amp; Ankle</td>
<td>10%</td>
</tr>
<tr>
<td>Hand &amp; Wrist</td>
<td>10%</td>
</tr>
<tr>
<td>Hip &amp; Knee</td>
<td>14%</td>
</tr>
<tr>
<td>Oncology</td>
<td>9%</td>
</tr>
<tr>
<td>Pediatrics</td>
<td>13%</td>
</tr>
<tr>
<td>Shoulder &amp; Elbow</td>
<td>10%</td>
</tr>
<tr>
<td>Spine</td>
<td>8%</td>
</tr>
<tr>
<td>Sports Medicine</td>
<td>5%</td>
</tr>
<tr>
<td>Trauma</td>
<td>12%</td>
</tr>
</tbody>
</table>

**Examination Overview**

**Subject Matter Experts**

Physician subject matter experts (SMEs) are appointed to the OITE Examination Committee by the Committee Appointment Program Committee, then ratified by the Board of Directors. The appointments cover a 3-year term and committee members may request reappointment near the end of their first term. SMEs are routinely surveyed.
regarding their focused expertise and every attempt is made to align a physician’s area of practice with appropriate content in the OITE. Based upon the content needs, assignments are issued to the SME groups. All SMEs receive training in item development focusing on the AAOS’ style for item writing as well as best practice design recognized by the standards in place for high-stakes and educational testing. Appendix A lists the SME contributors and reviewers for the 2021 OITE.

Item Development
SMEs author items in the AAOS item bank and authoring tool, ExamDeveloper®. Item writers complete their assignments in the tool and are required to provide detailed summaries of the topics covered in support of content validity. At least two current, relevant and peer reviewed/recognized sources of references are required for each test item as well as an in-depth discussion explaining the justification for the correct/best answer and explanation addressing the incorrect answer options. The extensive validation of each item provides the foundation for the content validity of the examination and supports the fundamental purpose of the OITE, considering the educational aspects of orthopaedic residency training.

Items are reviewed by the examination sub-committees prior to convening as a group to make decisions and recommendations regarding each item. The SME sub-committee members enter comments and suggestions for each item collection they are assigned to review. The group meets to address the collective comments and makes final edits to items worthy of retaining and including in the examination. On occasion, items are sent back to the authors for additional components (images, expanded discussion) and are reviewed upon completion of the requested additions. Each item is coded to a content domain and cognitive level and validated during the review process.

After the collection of items for each domain have undergone a series of reviews, a smaller committee (OITE chair and leads in each subject area domain) convene to assemble and approve the final form of the OITE. Currently, only one form is used across all PGY programs.

A similar development and review processes are followed for the collection of items shared by the ABOS Part I examination and the OITE. Each item is reviewed by a group of subject matter experts, revised as needed and coded to the appropriate content domain. The shared items included in the examination mapped to the OITE blueprint in consideration of the required overall compliance with content distribution.

Item Formats
OITE item format is traditionally four or five-part multiple-choice, with a single correct/best response key and three to four competing and compelling distractors (non-correct options). The items may present as short clinical cases (vignettes) or straightforward knowledge-based content. SMEs may include images, videos, or other stimuli (tables, charts, graphs) to supplement targeting the construct being tested. All
items are stand-alone, meaning they are not dependent on any other items in terms of sequence, association, or content.

Administration

In 2021, the OITE was available for administration under two models, the traditional, in-person, group testing and remote testing and proctoring. The OITE in-person, group administration is coordinated and proctored by each residency program. AAOS provided detailed instructions and guidelines to the residency programs regarding testing conditions, proctoring and requirements for the technical (hardware and bandwidth) components of the administration. The remote testing and proctoring administration were coordinated with a vendor that provided services for the test administration which included audio and video monitoring throughout the examination session. Residents were provided with instructions to download a secure browser in advance of the examination. In addition, a sample test was made available to all residents testing remotely a few weeks in advance of the OITE to ensure access was achieved to the server in advance of the live examination.

AAOS schedules a four-day window in November each year to cover the OITE administrations nationally and internationally. The testing window for the 2021 administration was November 5-14.

The examination is a 275-item multiple-choice item test and opens with a welcome from the Exams and Assessments Chair (see Appendix B). Residents/examinees may use up to 7 (seven) hours of testing time to complete the OITE. The examination is assembled in two sections. Residents have up to 3.5 hours of testing time to complete each section. During testing, examinees may flag items to review and are permitted to change answers. When finished with Section I, examinees are required to verify completion and submit their answers. Once completed, examinees are no longer permitted to access the test items to review or change answers. The same process is followed for the second section of the examination, with an opportunity to flag items to review, change answers, and verify completion of the examination. Prior to the delivery of the examination, examinees must review and accept the AAOS confidentiality/non-disclosure statement as well as participate in the pre-exam tutorial. The tutorial provides instruction on the layout, response options, and timing of the OITE. In addition, the tutorial displays how to enlarge images and play (replay) video clips.

Once the examination is complete, examinees are invited to respond to a brief survey relating to their testing experience.

Use of Scores

The OITE is a comprehensive examination designed to facilitate knowledge assessment in established principles and conventional procedures and treatment modalities in orthopaedic surgery. Orthopaedic residency programs use the OITE performance outcomes at the individual and programmatic levels to support orthopaedic resident education through study, research, discussion, review, and assessment.
Construct Irrelevant Variance

Fairness Review
During their training, SME item writers and reviewers are instructed to ensure that items are free from any detectable bias that could unduly advantage or disadvantage the test taker based upon individual characteristics. Careful attention to ensure that the construct being targeted is free from bias alleviates construct-irrelevant variance in the interpretation of scores.

Scoring Process

Item Analysis and Key Validation
Prior to scoring, an analysis of item performance and key validation study is conducted. Generally, items displaying negative discrimination (or point measure), proportion correct (p-value) of less than 0.20 or demonstrate evidence of a mis-key or double key are reviewed by the SMEs. Results data from the ACGME-accredited PGY 4 and PGY 5 testing cohorts were used to conduct the item analysis and key validation.

On occasion, an item may be determined to be mis-keyed and in that case, the item would be re-keyed and retained in the final scoring. Items that are determined to be flawed in any way (double key, not targeted, no longer relevant or accurate) are removed from scoring. For the 2021 OITE, 30 items were flagged for SME review and 12 items were removed from scoring for a final item count of N=263.

Classical and IRT Scoring

The AAOS scores the OITE using a classical scoring model (raw number correct) and reports the raw score, mean, standard deviation and percentile rank (individual and program). An extension of the scoring was conducted using a Rasch (item response theory) model to place the examinee performance outcomes and item difficulty measures on the same (logit) continuum. The Rasch model provides generalizability across samples and items, produces an ordered set of items, and identifies poorly functioning as well as unexpected responses. Examinees and items can be rank ordered while the interval between the measures is scalable and subject to richer interpretation than the classical model. 2,3

Results

Tables 2 and 3 show the overall scores by program and program year respectively
Table 2 Overall Results ACGME-accredited Programs and Non-ACGME

<table>
<thead>
<tr>
<th>Program Designation</th>
<th>N</th>
<th>Mean (SD)</th>
<th>% correct</th>
<th>Min Score</th>
<th>Max Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACGME</td>
<td>4466</td>
<td>166.09 (23.11)</td>
<td>63%</td>
<td>63</td>
<td>222</td>
</tr>
<tr>
<td>Non-ACGME</td>
<td>521</td>
<td>145.43 (23.18)</td>
<td>55%</td>
<td>62</td>
<td>202</td>
</tr>
</tbody>
</table>

Table 3 Overall Results by Program Year

<table>
<thead>
<tr>
<th>PGY</th>
<th>N</th>
<th>Mean (SD)</th>
<th>% correct</th>
<th>Min Score</th>
<th>Max Score</th>
<th>Mean Rasch Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGY 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACGME</td>
<td>881</td>
<td>185.54 (12.02)</td>
<td>71%</td>
<td>135</td>
<td>221</td>
<td>1.20</td>
</tr>
<tr>
<td>Non-ACGME</td>
<td>82</td>
<td>164.49 (20.22)</td>
<td>63%</td>
<td>68</td>
<td>202</td>
<td>0.73</td>
</tr>
<tr>
<td>PGY 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACGME</td>
<td>895</td>
<td>180.58 (13.31)</td>
<td>69%</td>
<td>63</td>
<td>222</td>
<td>1.08</td>
</tr>
<tr>
<td>Non-ACGME</td>
<td>120</td>
<td>154.61 (18.87)</td>
<td>59%</td>
<td>110</td>
<td>201</td>
<td>0.52</td>
</tr>
<tr>
<td>PGY 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACGME</td>
<td>900</td>
<td>172.28 (14.06)</td>
<td>66%</td>
<td>120</td>
<td>215</td>
<td>0.90</td>
</tr>
<tr>
<td>Non-ACGME</td>
<td>111</td>
<td>146.29 (19.38)</td>
<td>56%</td>
<td>62</td>
<td>189</td>
<td>0.34</td>
</tr>
<tr>
<td>PGY 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACGME</td>
<td>911</td>
<td>156.6 (15.54)</td>
<td>60%</td>
<td>91</td>
<td>205</td>
<td>0.56</td>
</tr>
<tr>
<td>Non-ACGME</td>
<td>121</td>
<td>134.28 (20.63)</td>
<td>51%</td>
<td>86</td>
<td>188</td>
<td>0.09</td>
</tr>
<tr>
<td>PGY 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACGME</td>
<td>879</td>
<td>135.34 (16.33)</td>
<td>51%</td>
<td>88</td>
<td>196</td>
<td>0.11</td>
</tr>
<tr>
<td>Non-ACGME</td>
<td>86</td>
<td>128.9 (19.83)</td>
<td>49%</td>
<td>94</td>
<td>184</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

ANOVA
An analysis of variance (ANOVA) was conducted to test the null hypothesis that there is no significant difference between mean scores across residency programs designated by program year. ACGME-accredited program data was used to conduct a comparison of mean scores across all PGY cohorts.

Results
The residency program year has significant influence on the overall OITE score, $F(4,4461)=1786.32$, $p=0.000$ (Table 4)
Table 4 ANOVA Results

ANOVA

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>1468723.686</td>
<td>4</td>
<td>367180.921</td>
<td>1786.321</td>
<td>.000</td>
</tr>
<tr>
<td>Within Groups</td>
<td>916965.309</td>
<td>4461</td>
<td>205.552</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2385688.994</td>
<td>4465</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5 displays the comparisons across all residency program years. Mean scores are significantly different between adjacent program years (PGY1->PGY2, PGY2->PGY3, PGY3->PGY4 and PGY4->PGY5). These results support evidence of an increase in knowledge acquisition and retention on the topics tested on in the OITE.

Table 5 Score Comparisons Across Adjacent PGY Cohorts

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>PGY (I)</th>
<th>PGY (J)</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig (p)</th>
<th>95% Confidence Interval</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tukey HSD</td>
<td>1</td>
<td>2</td>
<td>-21.255</td>
<td>0.678</td>
<td>0.000</td>
<td>-23.10 -19.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>-36.941</td>
<td>0.680</td>
<td>0.000</td>
<td>-38.80 -35.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3</td>
<td>-45.233</td>
<td>0.681</td>
<td>0.000</td>
<td>-47.09 -43.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4</td>
<td>-50.198</td>
<td>0.683</td>
<td>0.000</td>
<td>-52.06 -48.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>-15.686</td>
<td>0.674</td>
<td>0.000</td>
<td>-17.52 -13.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>5</td>
<td>-23.978</td>
<td>0.675</td>
<td>0.000</td>
<td>-25.82 -22.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>6</td>
<td>-28.943</td>
<td>0.677</td>
<td>0.000</td>
<td>-30.79 -27.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>-8.292</td>
<td>0.677</td>
<td>0.000</td>
<td>-10.14 -6.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>6</td>
<td>-13.257</td>
<td>0.679</td>
<td>0.000</td>
<td>-15.11 -11.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>5</td>
<td>-4.965</td>
<td>0.680</td>
<td>0.000</td>
<td>-6.82 -3.11</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Item Resident Distribution
The following figures show the distribution of examinee ability estimates and item difficulty estimates on the Rasch interval scale for the PGY cohorts. The mean item difficulty is centered at zero. Measures to the left (negative) represent measures that align with “easier” items with the easiest items at the far left of the continuum. Measures to the right of the mean, represent items that increase in difficulty for the residents as the scale moves further right. The distributions show the alignment between candidate ability and item difficulty. There are a number of items that tested very easy (left end of lower scale) across all cohorts. These items contribute minimally to measurement precision (all or nearly all test-takers scored correct) and likely influence reliability estimates (see Reliability). The diagrams are used as a component of exam review and are not intended to suggest any intentional effort to increase the difficulty of the test items.

Figure 1 PGY 5

Figure 2 PGY 4
Reliability
Test form reliability is a measure of internal consistency of the examination and more importantly the magnitude of measurement error. Reliability measures are typically reported on a scale from 0-1. In educational assessment, reliability may be interpreted as the confidence that the scores accurately and consistently measure the knowledge of the test-takers.

KR20 reliability measures the repeatability of raw scores and may be affected by the overall test difficulty, number of items on the test, items that did not discriminate and the spread of scores. The Rasch Person Reliability Index produces a measure indicating how repeatable the measures are for the sample. There is good evidence to show the reproducibility of outcomes if the same set of items were tested in similarly able populations of examinees. The reliability estimates are acceptable for an educational examination.

The standard error of measurement (SEM) estimates the variation in test scores within a sample of test-takers. Table 6 shows the reliability estimates and SEM across the five testing cohorts.

Table 6 Reliability Estimates (Overall exam) of the 2021 OITE Across Program Years

<table>
<thead>
<tr>
<th></th>
<th>KR20</th>
<th>SEM raw score</th>
<th>Rasch Person Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGY 1</td>
<td>0.81</td>
<td>7.06</td>
<td>0.83</td>
</tr>
<tr>
<td>PGY 2</td>
<td>0.80</td>
<td>6.89</td>
<td>0.81</td>
</tr>
<tr>
<td>PGY 3</td>
<td>0.78</td>
<td>6.60</td>
<td>0.77</td>
</tr>
<tr>
<td>PGY 4</td>
<td>0.77</td>
<td>6.40</td>
<td>0.75</td>
</tr>
<tr>
<td>PGY 5</td>
<td>0.73</td>
<td>6.27</td>
<td>0.71</td>
</tr>
</tbody>
</table>

ABOS/AAOS Linking Study
The American Board of Orthopaedic Surgery (ABOS) and the American Academy of Orthopaedic Surgeons (AAOS) collaborated on the development of a collection of examination items (questions) that were included on both the 2021 ABOS Part I Certifying Examination and the 2021 AAOS Orthopaedic In-Training Examination (OITE).

The purpose of including a set of common items on both examinations was to identify the score on the 2021 AAOS OITE that approximately corresponds to the minimum passing performance level on the ABOS Part I Certifying Examination. This approximation is based on a relatively small sample of shared items and is not a guarantee of or predictive of future performance on the ABOS Part I Examination.

Based on the linking study, the score on the 2021 AAOS OITE that corresponds to the ABOS Part I minimal passing standard is 182 items or 69.2% correct. Many factors, including
changing levels of knowledge and testing conditions, will impact one’s performance on the AAOS OITE and the ABOS Part I Examinations. This information is provided as a rough benchmark to help guide education and examination preparations. The number correct score and corresponding percent correct are applicable to the 2021 AAOS OITE only and should not be used to gauge performance on previous or future administrations of the OITE.

Linking Study Design
The shared items developed by ABOS and AAOS were pre-tested on the ABOS Part I Examination in July 2021. The item performance measures for each item were analyzed and reviewed by subject matter experts who made the final decision regarding which items would be included on the 2021 AAOS OITE. Items that did not meet statistical specifications were eliminated from the linking study.

Responses to these shared items from PGY 5 residents in US residency programs who took the 2021 AAOS OITE and from candidates who took the 2021 ABOS Part I Certifying Examination were used in a linking study—a psychometric procedure to link two examinations that have items in common. Item calibrations for the shared items produced from the ABOS Part I Examination were used to estimate item measures on the AAOS OITE. The current passing standard for the ABOS Part I Examination was applied to the OITE final item measures to obtain an estimate of the number correct and corresponding percentage correct that were reflective of the ABOS Part I Examination minimum passing score.

Results
The results of applying the minimal passing standard to the ACGME-accredited PGY cohorts demonstrated remarkable findings in terms of performance.

Figure 6 Percentage of Scores Meeting or Exceeding the ABOS Part I Minimum Passing Standard

<table>
<thead>
<tr>
<th>PGY Level</th>
<th>Percentage Meeting or Exceeding ABOS Minimum Passing Standard Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGY 5</td>
<td>65.8%</td>
</tr>
<tr>
<td>PGY 4</td>
<td>50.1%</td>
</tr>
<tr>
<td>PGY 3</td>
<td>25.6%</td>
</tr>
<tr>
<td>PGY 2</td>
<td>6.0%</td>
</tr>
<tr>
<td>PGY 1</td>
<td>0.3%</td>
</tr>
</tbody>
</table>
Recommendations
The following recommendations address test development and measurement considerations.

1. Review and revise (as necessary) the purpose of the OITE. It is important for test developers to fully embrace the fundamental purpose of an examination and execute the test development processes and test specifications to align with said purpose.

2. Review post-examination comments to support quality improvement

3. Equate adjacent administration forms of the OITE. Under the current model of test form assembly, all items are released for public review. Not only does this contribute to an extensive amount of item development necessary from year to year, it prohibits the ability to equate the test forms with a common item set. Currently, the OITE form assembly does not consider any statistical parameters (i.e. form difficulty). Equating the OITE will provide for better interpretability of scores from year to year (See Test Equating Proposal).

4. Continue to expand the groups of SMEs who contribute to the item development process. As committee members become seasoned with the AAOS style and process of item development and review, they may be used in a greater capacity as mentors and reviewers.

5. Implement item response theory (IRT), Rasch scoring in addition to classical scoring. This scoring model compliments test equating and produces meaningful measurement of item and examinee performance on an interval scale. The Rasch model also provides generalizability across samples and items, produces an ordered set of items, and identifies poorly functioning as well as unexpected responses.

6. Evaluate item and person performance at the cohort level to help guide future item development. There are a significant number of items that tested extremely easy across all groups. This may be essential content that must be tested and thus it is of paramount importance to include the recommendations of the SMEs before eliminating items based solely on statistical outcomes.

7. Given the positive outcome from the dual model administration, the AAOS should offer the OITE in both testing formats for future administrations.
Summary
The 2021 OITE test development and analyses process align consistently with the standards set forth for professional, high-stakes examinations. The steps to define the purpose of the examination and use of scores, selection of SMEs, training, item development, review and validation follow rigorous standards set forth by the AAOS Education Assessments and Examinations Committee members.

The 2021 OITE provides key information relating to the progression of knowledge in the topics covered for orthopaedic residents across their educational trajectory. Educational exams by design should differentiate among cohorts that are categorized or distinguished by their place in the program. The 2021 OITE outcomes show that the examination measures what it is purported to measure as evidenced by the differences in scores among residents in training over time. The 2021 OITE data supports the expectation of growth over time by the presentation of composite scores that are significantly different across the PGY cohorts.

The significant efforts by the volunteer physicians whose contributions support the overall positive results and commendable outcomes of the 2021 OITE should be recognized for their essential role in the OITE test development process.
Reference List


Appendix A
2021 Education Assessments and Examinations
Committee and Contributors

Norman Y. Otsuka, MD, FAAOS, Editor in Chief

Professor and Chair, Division of Orthopaedic Surgery, Southern Illinois University School of Medicine, Springfield, Illinois

Contributors

D. Gordon Allan, MD, FAAOS
Reconstructive Orthopedic Surgery, Private Practice, Orthopedic Center of Illinois, Springfield, Illinois

*Craig S. Bartlett, MD, FAAOS
Professor of Orthopaedics, Medical Director of Orthopaedic Trauma, University of Vermont, Burlington, Vermont

*Claire F. Beimesch, MD, FAAOS
Pediatric Orthopaedic Surgeon, Dayton Children’s Hospital, Dayton, Ohio

*Yelena Bogdan, MD, FACS, FAAOS
Orthopaedic Trauma Surgeon, Director of Orthopaedic Education, NYC Health and Hospitals/Jacobi, Assistant Professor of Orthopaedic Surgery, Albert Einstein College of Medicine, Bronx, New York

*Brian E. Brigman, MD, PhD, FAAOS
Associate Professor of Orthopaedic Surgery and Pediatrics, Chief, Section of Orthopaedic Oncology, Sarcoma Program Director, Duke Cancer Institute, Associate Program Director, Duke Orthopaedic Residency Duke University, Durham, North Carolina

*Robert H. Cho, MD, FAAOS
Chief of Staff, Shriners For Children Medical Center, Pasadena, California; Clinical Assistant Professor,
University of California Los Angeles, Department of Orthopaedic Surgery, Los Angeles, California

*Lawrence F. Cohen, MD, FAAOS
Associate Professor, Department of Orthopedic Surgery, Virginia Commonwealth University Health, Richmond, Virginia
*Ilvy Cotterell, MD, FAAOS

Associate Professor, Department of Orthopaedic Surgery, Virginia Commonwealth University, Chief, Division of Orthopaedic Surgery, Hunter Holmes McGuire Veterans Affairs Medical Center, Richmond, Virginia

*Thomas C. Dowd, MD, FAAOS

Program Director, San Antonio Uniformed Services Health Education Consortium, Orthopaedic Residency, San Antonio, Texas; Associate Professor, Uniformed Services University of the Health Sciences, Bethesda, Maryland

*Patrick B. Ebeling, MD, FAAOS

Faculty, Twin Cities Foot and Ankle Fellowship, Eagan, Minnesota; Orthopaedic Foot and Ankle Surgeon, Private Practice, Twin Cities Orthopedics, Burnsville, Minnesota

*David S. Geller, MD, FAAOS

Associate Professor, Department of Orthopaedic Surgery and Pediatrics, Albert Einstein College of Medicine; Vice-Chairman of Strategy and Innovation, Department of Orthopaedic Surgery, Montefiore Medical Center; Co-Director, Orthopaedic Oncology Division, Montefiore Medical Center, Bronx, New York

*Brian M. Grawe, MD, FAAOS

Associate Professor, Sports Medicine & Shoulder Reconstruction, Department of Orthopaedic Surgery, University of Cincinnati Academic Health Center, Cincinnati, Ohio

*Chancellor F. Gray, MD, FAAOS

Associate Professor, University of Florida College of Medicine, Department of Orthopaedic Surgery, Gainesville, Florida

*Brian M. Haus, MD, FAAOS

Associate Professor of Clinical Orthopaedic Surgery, University of California, Davis Health, Chief, Pediatric Orthopaedic Surgery, Department of Orthopaedic Surgery, UC Davis Children’s Hospital; Associate Professor of Clinical Orthopaedic Surgery, Shriners Hospital for Children, (joint appointment with UC Davis); Co-Director, Pediatric Sports Medicine and Joint Preservation Shriners Hospital for Children, Northern California, Sacramento, California

*John W. Hinchey, MD, FAAOS

Orthopaedic Shoulder & Elbow Surgeon, Private Practice, Ortho San Antonio; Adjunct Assistant Professor, Department of Orthopaedic Surgery, UT Health San Antonio, Assistant Fellowship Director, Rockwood Shoulder and Elbow Fellowship, UT Health San Antonio, San Antonio, Texas
*J. Gabriel Horneff III, MD, FAAOS  
Assistant Professor of Clinical Orthopaedics, Department of Orthopaedic Surgery, University of Pennsylvania, Philadelphia, Pennsylvania

*J. Benjamin Jackson III, MD, MBA, FAOA, FACS, FAAOS  
Associate Professor, University of South Carolina, Department of Orthopaedics, Director of Orthopaedic Research, Medical Director, Foot and Ankle Division, Charleston, South Carolina

*Marci D. Jones, MD, FAAOS  
Associate Professor, Vice Chair (Education), Director of Hand Fellowship, Department of Orthopedic Surgery, University of Massachusetts, Worcester, Massachusetts

*Kyle T. Judd, MS, MD, FACS, FAAOS  
Assistant Professor, Fellowship Director of Orthopaedic Trauma, University of Rochester, Department of Orthopaedic Surgery, Division of Trauma, Rochester, New York

*Raymond A. Liu, MD, FAAOS  
Associate Professor of Orthopedics and Pediatrics, Case Western Reserve University School of Medicine; Pediatric Orthopedic Surgeon, University Hospitals Rainbow Babies & Children’s, and the Victor M. Goldberg Endowed Chair in Orthopedics, Cleveland, Ohio

*David V. Lopez, MD, FAAOS  
Assistant Professor, Department of Orthopedic Surgery, Hackensack Meridian School of Medicine, Nutley, New Jersey; Orthopaedic Sports Medicine Surgeon, Private Practice, Orthopaedic and Sports Medicine Specialists, Little Silver, New Jersey

*Richard L. McGough, III, MD, FAAOS  
Chief, Division of Musculoskeletal Oncology, Professor, Department of Orthopaedic Surgery, Professor, Department of Surgery (Surgical Oncology), School of Medicine, Professor, Department of Rehabilitation Science and Technology, School of Health and Rehabilitation Sciences, University of Pittsburgh; Co-Director, Sarcoma Services, UPMC Cancer Centers, Pittsburgh, Pennsylvania

*Kathleen A. McHale, MD, FAAOS  
Professor of Surgery at the Uniform Services University for Health Sciences, Bethesda, Maryland; Clinical Professor of Orthopaedic Surgery at George Washington University, Washington, DC; Clinical Professor of Orthopaedic Surgery, Southern Illinois University, Springfield, Illinois
*Varatharaj Mounasamy, MD, FAAOS
Professor, Department of Orthopedic Surgery, UT Southwestern Medical Center, Section Chief Orthopaedic Surgery, VA Medical Center, Dallas, Texas

*Stephanie J. Muh, MD, FAAOS
Henry Ford Hospital Department of Orthopaedic Surgery, Deputy Chief, Orthopaedic Surgery Service, Henry Ford West Bloomfield Hospital, Associate Program Director - Sports Medicine Fellowship, Detroit, Michigan

Matthew E. Oetgen, MD, MBA, FAAOS
Chief, Division of Orthopaedic Surgery and Sports Medicine, Children's National Hospital, Associate Professor, Department of Orthopaedic Surgery, George Washington University, Washington, DC

*Shervin V. Oskouei, MD, FAAOS
Assistant Professor, Department of Orthopaedic Surgery, Emory University School of Medicine, Chief of Orthopedic Surgery, Emory University Hospital Midtown, Division Director of Orthopaedic Oncology, Atlanta, Georgia

*Jonathan D. Packer, MD, FAAOS
Associate Professor, Department of Orthopaedics, University of Maryland School of Medicine, Baltimore, Maryland

Stephen J. Pineda, MD, FAAOS
Assistant Clinical Professor of Orthopedic Surgery, Southern Illinois University School of Medicine, Springfield, Illinois

*Greg W. Rennirt, MD, FAAOS
Orthopedic Sports Medicine Surgeon, Private Practice, Louisville Kentucky

*Milan K. Sen, MD, FAAOS
Associate Professor, Department of Orthopaedic Surgery, Albert Einstein College of Medicine; Chief of Orthopedic Surgery, NYC Health+Hospitals/Jacobi, Director of Orthopedic Trauma, NYC Health+Hospitals/Jacobi, Bronx, New York

*Kenneth F. Taylor, MD, FAAOS
Chief, Division of Hand Surgery, Professor, Department of Orthopaedic Surgery and Rehabilitation, Penn State College of Medicine, Penn State Health Bone and Joint Institute, Hershey, Pennsylvania

*Paul Toogood, MD, FAAOS
Assistant Professor of Orthopedics, University of California San Francisco, San Francisco, California
*Krishna R. Tripuraneni, MD, FAAOS
Orthopaedic Surgeon, Hip and Knee Arthroplasty, New Mexico Orthopaedic Associates, Albuquerque, New Mexico

*Harmeeth S. Uppal, MD, MS, FACS, FAAOS
Orthopaedic Trauma Surgeon, Private Practice, Orthopaedic Trauma and Fracture Specialists, San Diego, California

*Brian C. Werner, MD, FAAOS
Assistant Professor, Department of Orthopaedic Surgery, University of Virginia, Charlottesville, Virginia; Head Team Physician, James Madison University, Harrisonburg, Virginia

*Nadine L. Williams, MD
Assistant Professor, Department of Orthopaedic Surgery, Loma Linda University Medical Center, Loma Linda, California

Christopher E. Wottowa, MD, FAAOS
Orthopaedic Shoulder & Elbow Surgeon, Private Practice, Passavant Area Hospital, Jacksonville, Illinois; Memorial Medical Center, Springfield, Illinois; HSHS St. John’s Hospital, Springfield, Illinois; Abraham Lincoln Memorial Hospital, Lincoln, Illinois; Ambulatory Surgery & Endoscopy Center, Springfield, Illinois

*John S. Xenos, MD, FAAOS
Fellowship Director, Colorado Orthopaedic Research Institute, President, Orthopaedic Adult Reconstruction Surgeon, Colorado Orthopaedics, Lone Tree, Colorado

Yi-Meng Yen, MD, PhD, FAAOS
Assistant Professor, Children’s Hospital Boston, Department of Orthopaedic Surgery, Division of Sports Medicine, Child and Young Adult Hip Preservation Program, Boston, Massachusetts

*Denotes Committee Appointment
Staff Contributors
Irene Bogdal, Examinations Administrative Specialist
Christina Fulgaro Crumlish, MA - Examinations Specialist
Marcie Lampert - Senior Examinations Specialist
David Marshall, CPTM- Manager Assessments and Examinations
November 2021

Dear Residents and Colleagues,

On behalf of the Assessments and Examinations Committee, I would like to thank you for participating in this year’s examination.

Being the first such examination among all the medical specialties, this year marks the 57th administration of the Orthopaedic In-training Examination®.

Every year, the OITE® is created by a group of orthopaedic surgeons who are deeply committed to education through a rigorous peer-review process. While many topics in orthopaedic surgery lack conclusive evidence, many of our practices are based on sound medical principles and a consensus of experts. Sometimes, there may appear several correct answers to a test item, but each question has been vetted through this peer-review process to select the one best preferred response.

With this and each subsequent administration, we strive to make the OITE® better. Thank you for being part of this year’s examination, and I hope that you find this to be an enriching educational experience.

Sincerely,

[Signature]

Norman Otsuka, MD
Chair, Assessments and Examinations Committee