

**The PAEA End of Rotation™ Exams:
Historical Statistics From 2013 to 2019**



**The PAEA End of Rotation™ Exams:
Historical Statistics From 2013 to 2019**

Published March 2023

Prepared by PAEA Assessment

Introduction

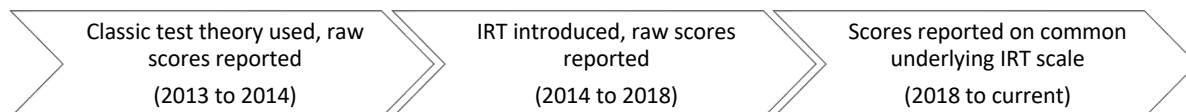
The PAEA End of Rotation exams are a set of objective, standardized evaluations intended to serve as one measure of the medical knowledge students gain during specific supervised clinical practice experiences.

The PAEA End of Rotation scoring, and reporting methods have evolved over time. Traditionally, the projected national comparative data for the current exam forms (i.e., the expected performance on the *current exam forms* are based on a representative sample of PA students from *previous exams*) has been housed on the PAEA webpage, along with the final observed statistics for previous versions. Moving forward, the national comparative data for the current set of exam forms will remain on the website for ease of access. Historical statistics, for retired (i.e., no longer active) exam forms, will be stored in this document for reference. The following sections describe how End of Rotation score reporting has changed over time and helps to provide context for how scores were reported at the time of exam delivery. Tables containing the final statistics for retired exam forms follow the subsequent explanatory sections.

Scoring and Reporting Evolution

Version 1 of the PAEA End of Rotation exams, published in 2013, were initially developed and scored using the measurement model of classical test theory. At the time, classical test theory was appropriate given the number of exam forms published and students taking the exams (sometimes described as exam volume). As the item banks and student volume grew, exam development and scoring for version 2 of the End of Rotation exams, published in 2014, were migrated to item response theory

(IRT). The use of IRT provided advantages over classical methods, specifically in building parallel test forms and relating the difficulty of exam items to a single underlying metric. With the release of version 6 of the End of Rotation exams in 2018, PAEA moved to reporting scale scores based on the underlying IRT metric¹ rather than the number of correct responses (i.e., raw scores) typical of non-equated classical test theory programs. In evaluating or comparing student performance, the use of scale scores (or transformations of them, such as percentiles or z-scores) is preferable to raw scores. While End of Rotation exam forms are built to be similar in difficulty, raw scores are still susceptible to differences in difficulty between forms, while scale scores are not. Regardless of the reporting metric, comparative national data has been provided to help programs benchmark student and program performance since 2013.²



National Comparative Data

One of the requirements to ensure comprehensive utility of the PAEA End of Rotation exams is the reporting of national comparative data to PA programs and students. This is important for two main reasons:

1. It provides a way to benchmark the students' knowledge relative to a larger population of students.

¹ Note, while version 6 of the End of Rotation exams were the first to be built with the scale scores in mind, version 5 scores were put on the scale score metric and published retroactively for reference/comparison.

² The only instance in which comparative data was not available was in the early stages of version 1, due to relatively small initial sample sizes and thus unstable statistics. Once sample sizes of 200 student test-takers were achieved in each exam program, national data was provided to aide normative interpretation of scores.

2. It provides PA programs with valuable information that can potentially help inform decision-making regarding curriculum or clinical experiences for content areas where their students performed lower than the mean.

Like the End of Rotation scoring methods, the way in which national comparative data has been calculated has shifted since the inception of the program. The only time national comparative data was not provided to PA programs was in the first nine (9) months of the End of Rotation program (i.e., January to mid-October of 2013), during which time it was necessary to delay the release of aggregated results until appropriate sample sizes could be achieved to support robust statistics. Once such a sample size was achieved, national statistics were published and served as points of comparison for students and PA programs.

With the release of version 2 of the End of Rotation exams, a base reference group approach was implemented, which allowed comparative statistics to be released simultaneously with the publication and administration of new exam forms. From version 2 onward, and regardless of whether a classical or IRT measurement model was in use at the time (see above), the general approach was to provide expected performance statistics for the current exam forms based on how students had previously performed on the items that made up those forms.

COVID-19 Pandemic Impact

Version 7 of the PAEA End of Rotation exams was published in July of 2019. Due to the extenuating circumstances of the COVID-19 pandemic and its impact on assessment efforts by PA programs, version 7 forms remained available for administration for an extended period. Additionally, standard proctoring procedures that had been in place were waived to allow programs more flexibility in assessing their students. PAEA closely monitored student performance during this time to

evaluate the impact changes to proctoring protocols and the extended publication time of version 7 may have on nationally aggregated statistics. The results of these analyses indicated no meaningful increase in the End of Rotation scores relative to the historical data. Given these findings, programs may still compare scores, particularly scale scores which account for the difficulty of the items comprising a form, across years and before, during, and after the pandemic. That said, and while no meaningful shift in scores was observed at the national level, programs were encouraged to be mindful of the unprecedented circumstances the pandemic brought to bear on PA programs and students when considering the performance of their own program during this time.

Emergency Medicine

Raw Score Table (score range of 0-100)

Administration Dates	Form	Mean	SD	Reliability	SEM
1/15/2013 - 7/15/2014	1.1	70.1	8.4	Not reported	Not reported
1/15/2013 - 7/15/2014	1.2	69.2	9.1	Not reported	Not reported
7/16/2014 - 7/7/2015	2.1	73.7	8.4	0.78	4.0
7/16/2014 - 7/7/2015	2.2	73.0	8.7	0.79	4.0
7/8/2015 - 7/7/2016	3.1	75.5	9.0	0.81	3.9
7/8/2015 - 7/7/2016	3.2	74.7	8.8	0.79	4.1
7/8/2016 - 7/7/2017	4.1	75.4	8.4	0.78	3.9
7/8/2016 - 7/7/2017	4.2	75.8	8.4	0.78	3.9
7/8/2017 - 7/10/2018	5.1	77.4	9.1	0.81	4.0
7/8/2017 - 7/10/2018	5.2	77.1	9.0	0.81	3.9
7/11/2018 - 7/16/2019	6.1	78.8	9.0	0.81	3.8
7/11/2018 - 7/16/2019	6.2	77.6	8.8	0.81	3.9

Emergency Medicine

Scale Score Table (score range of 300-500)

Administration Dates	Form	Mean	SD	Reliability	SEM
7/8/2017 - 7/10/2018	5.1, 5.2	405 - 406	20 - 21	0.81	9
7/11/2018 - 7/16/2019	6.1, 6.2	410	21	0.81	9
7/17/2019 - Current Coming Soon					

Note regarding Forms 5 and 6: Despite scale scores not being reported until version 6 of End of Rotation exams, scale score statistics for version 5 of the exams were retroactively calculated and released to assist with interpretation and help programs who may have had students testing on both sides of the version 5 to version 6 migration.

Family Medicine

Raw Score Table (score range of 0-100)

Administration Dates	Form	Mean	SD	Reliability	SEM
1/15/2013 - 7/15/2014	1.1	75.1	7.7	Not reported	Not reported
1/15/2013 - 7/15/2014	1.2	75.8	8.0	Not reported	Not reported
7/16/2014 - 7/7/2015	2.1	75.3	8.1	0.76	3.9
7/16/2014 - 7/7/2015	2.2	75.4	8.0	0.75	4.0
7/8/2015 - 7/7/2016	3.1	77.0	8.8	0.80	4.1
7/8/2015 - 7/7/2016	3.2	77.1	8.6	0.79	4.0
7/8/2016 - 7/7/2017	4.1	77.7	8.0	0.77	3.8
7/8/2016 - 7/7/2017	4.2	78.5	8.1	0.78	3.8
7/8/2017 - 7/10/2018	5.1	77.9	8.7	0.80	3.8
7/8/2017 - 7/10/2018	5.2	78.9	8.5	0.80	3.8
7/8/2017 - 7/10/2018	5.3	78.8	8.3	0.79	3.8
7/11/2018 - 7/16/2019	6.1	78.4	8.6	0.80	3.9
7/11/2018 - 7/16/2019	6.2	79.6	8.3	0.80	3.8
7/11/2018 - 7/16/2019	6.3	79.6	8.6	0.80	3.8

Family Medicine

Scale Score Table (score range of 300-500)

Administration Dates	Form	Mean	SD	Reliability	SEM
7/8/2017 - 7/10/2018	5.1, 5.2, 5.3	401 - 406	25 - 26	0.79 - 0.80	11 - 12
7/11/2018 - 7/16/2019	6.1, 6.2, 6.3	407	26	0.80	12
7/17/2019 - Current Coming Soon					

Note regarding Forms 5 and 6: Despite scale scores not being reported until version 6 of End of Rotation exams, scale score statistics for version 5 of the exams were retroactively calculated and released to assist with interpretation and help programs who may have had students testing on both sides of the version 5 to version 6 migration.

General Surgery

Raw Score Table (score range of 0-100)

Administration Dates	Form	Mean	SD	Reliability	SEM
1/15/2013 - 7/15/2014	1.1	71.3	8.9	Not reported	Not reported
1/15/2013 - 7/15/2014	1.2	72.4	8.1	Not reported	Not reported
7/16/2014 - 7/7/2015	2.1	72.7	9.3	0.81	4.0
7/16/2014 - 7/7/2015	2.2	75.1	9.0	0.81	3.9
7/8/2015 - 7/7/2016	3.1	75.4	9.4	0.82	4.0
7/8/2015 - 7/7/2016	3.2	76.2	8.8	0.81	3.9
7/8/2016 - 7/7/2017	4.1	77.4	8.6	0.80	3.8
7/8/2016 - 7/7/2017	4.2	76.8	9.0	0.82	3.8
7/8/2017 - 7/10/2018	5.1	76.2	8.4	0.78	3.9
7/8/2017 - 7/10/2018	5.2	76.2	8.4	0.78	3.9
7/11/2018 - 7/16/2019	6.1	78.3	8.0	0.77	3.8
7/11/2018 - 7/16/2019	6.2	77.7	8.3	0.77	4.0

General Surgery

Scale Score Table (score range of 300-500)

Administration Dates	Form	Mean	SD	Reliability	SEM
7/8/2017 - 7/10/2018	5.1, 5.2	403 - 404	23, 24	0.77 - 0.78	11
7/11/2018 - 7/16/2019	6.1, 6.2	410	24	0.77	12
7/17/2019 - Current Coming Soon					

Note regarding Forms 5 and 6: Despite scale scores not being reported until version 6 of End of Rotation exams, scale score statistics for version 5 of the exams were retroactively calculated and released to assist with interpretation and help programs who may have had students testing on both sides of the version 5 to version 6 migration.

Internal Medicine

Raw Score Table (score range of 0-100)

Administration Dates	Form	Mean	SD	Reliability	SEM
1/15/2013 - 7/15/2014	1.1	68.2	9.7	Not reported	Not reported
1/15/2013 - 7/15/2014	1.2	76.7	8.2	Not reported	Not reported
7/16/2014 - 7/7/2015	2.1	74.8	7.9	0.75	3.9
7/16/2014 - 7/7/2015	2.2	73.4	7.7	0.74	3.9
7/8/2015 - 7/7/2016	3.1	74.4	8.6	0.78	4.1
7/8/2015 - 7/7/2016	3.2	75.2	8.1	0.76	4.1
7/8/2016 - 7/7/2017	4.1	74.5	8.1	0.77	3.9
7/8/2016 - 7/7/2017	4.2	75.2	7.9	0.76	3.9
7/8/2017 - 7/10/2018	5.1	77.2	9.2	0.82	3.9
7/8/2017 - 7/10/2018	5.2	77.0	8.7	0.80	3.9
7/11/2018 - 7/16/2019	6.1	78.9	8.9	0.82	3.9
7/11/2018 - 7/16/2019	6.2	78.5	9.2	0.82	3.9

Internal Medicine

Scale Score Table (score range of 300-500)

Administration Dates	Form	Mean	SD	Reliability	SEM
7/8/2017 - 7/10/2018	5.1, 5.2	404 - 406	23 - 24	0.80 - 0.82	10
7/11/2018 - 7/16/2019	6.1, 6.2	409	24	0.82	10
7/17/2019 - Current Coming Soon					

Note regarding Forms 5 and 6: Despite scale scores not being reported until version 6 of End of Rotation exams, scale score statistics for version 5 of the exams were retroactively calculated and released to assist with interpretation and help programs who may have had students testing on both sides of the version 5 to version 6 migration.

Pediatrics

Raw Score Table (score range of 0-100)

Administration Dates	Form	Mean	SD	Reliability	SEM
1/15/2013 - 7/15/2014	1.1	72.6	7.4	Not reported	Not reported
1/15/2013 - 7/15/2014	1.2	75.1	7.8	Not reported	Not reported
7/16/2014 - 7/7/2015	2.1	74.4	7.9	0.76	3.86
7/16/2014 - 7/7/2015	2.2	75.7	7.8	0.76	3.85
7/8/2015 - 7/7/2016	3.1	76.5	8.5	0.79	3.96
7/8/2015 - 7/7/2016	3.2	76.0	8.5	0.79	4.05
7/8/2016 - 7/7/2017	4.1	76.5	8.5	0.78	3.97
7/8/2016 - 7/7/2017	4.2	76.8	8.7	0.80	3.95
7/8/2017 - 7/10/2018	5.1	77.3	8.7	0.80	3.87
7/8/2017 - 7/10/2018	5.2	77.8	8.3	0.80	3.72
7/11/2018 - 7/16/2019	6.1	80.7	7.9	0.79	3.69
7/11/2018 - 7/16/2019	6.2	80.3	8.3	0.79	3.79

Pediatrics

Scale Score Table (score range of 300-500)

Administration Dates	Form	Mean	SD	Reliability	SEM
7/8/2017 - 7/10/2018	5.1, 5.2	405 - 406	25	0.80	11
7/11/2018 - 7/16/2019	6.1, 6.2	412	25	0.79	12
7/17/2019 - Current Coming Soon					

Note regarding Forms 5 and 6: Despite scale scores not being reported until version 6 of End of Rotation exams, scale score statistics for version 5 of the exams were retroactively calculated and released to assist with interpretation and help programs who may have had students testing on both sides of the version 5 to version 6 migration.

Psychiatry and Behavioral Health

Raw Score Table (score range of 0-100)

Administration Dates	Form	Mean	SD	Reliability	SEM
1/15/2013 - 7/15/2014	1.1	75.6	6.7	Not reported	Not reported
1/15/2013 - 7/15/2014	1.2	73.0	6.5	Not reported	Not reported
7/16/2014 - 7/7/2015	2.1	77.7	7.8	0.78	3.7
7/16/2014 - 7/7/2015	2.2	80.9	6.8	0.74	3.5
7/8/2015 - 7/7/2016	3.1	79.3	7.6	0.77	3.9
7/8/2015 - 7/7/2016	3.2	78.3	7.8	0.77	3.8
7/8/2016 - 7/7/2017	4.1	82.6	6.9	0.74	3.5
7/8/2016 - 7/7/2017	4.2	82.7	6.7	0.74	3.4
7/8/2017 - 7/10/2018	5.1	84.3	6.6	0.74	3.4
7/8/2017 - 7/10/2018	5.2	83.3	6.5	0.72	3.4
7/11/2018 - 7/16/2019	6.1	83.7	6.4	0.72	3.4
7/11/2018 - 7/16/2019	6.2	82.7	6.6	0.72	3.5

Psychiatry and Behavioral Health

Scale Score Table (score range of 300-500)

Administration Dates	Form	Mean	SD	Reliability	SEM
7/8/2017 - 7/10/2018	5.1, 5.2	404 - 406	23 - 25	0.72 - 0.74	12 - 13
7/11/2018 - 7/16/2019	6.1, 6.2	412	24	0.72	13
7/17/2019 - Current Coming Soon					

Note regarding Forms 5 and 6: Despite scale scores not being reported until version 6 of End of Rotation exams, scale score statistics for version 5 of the exams were retroactively calculated and released to assist with interpretation and help programs who may have had students testing on both sides of the version 5 to version 6 migration.

Women's Health

Raw Score Table (score range of 0-100)

Administration Dates	Form	Mean	SD	Reliability	SEM
1/15/2013 - 7/15/2014	1.1	72.6	7.4	Not reported	Not reported
1/15/2013 - 7/15/2014	1.2	73.3	7.0	Not reported	Not reported
7/16/2014 - 7/7/2015	2.1	74.3	7.3	0.72	3.84
7/16/2014 - 7/7/2015	2.2	74.8	6.7	0.70	3.71
7/8/2015 - 7/7/2016	3.1	75.4	8.4	0.78	4.04
7/8/2015 - 7/7/2016	3.2	74.9	8.2	0.76	4.10
7/8/2016 - 7/7/2017	4.1	75.5	8.1	0.78	3.84
7/8/2016 - 7/7/2017	4.2	74.9	7.6	0.74	3.87
7/8/2017 - 7/10/2018	5.1	75.9	8.3	0.76	4.05
7/8/2017 - 7/10/2018	5.2	76.4	8.7	0.79	4.01
7/11/2018 - 7/16/2019	6.1	77.3	8.3	0.79	3.9
7/11/2018 - 7/16/2019	6.2	75.7	8.8	0.79	4.0

Women's Health

Scale Score Table (score range of 300-500)

Administration Dates	Form	Mean	SD	Reliability	SEM
7/8/2017 - 7/10/2018	5.1, 5.2	400 - 403	24 - 25	0.76 - 0.79	11 - 12
7/11/2018 - 7/16/2019	6.1, 6.2	405	24	0.79	11
7/17/2019 - Current Coming Soon					

Note regarding Forms 5 and 6: Despite scale scores not being reported until version 6 of End of Rotation exams, scale score statistics for version 5 of the exams were retroactively calculated and released to assist with interpretation and help programs who may have had students testing on both sides of the version 5 to version 6 migration.