The testing plan below was used during the Testing and Learning Phase of the Measure Development Lifecycle for the NEMSQA EMS Compass 2.0 Measure Re-Specification Project. Plan was developed and executed June-July 2019.

Measure Testing Plan

I. Task Objective: To analyze revised EMS Compass measures for validity and reliability.

II. Approach:

The Testing & Learning Workgroup will evaluate select scientific properties of EMS Compass measures that have been re-specified by NEMSQA. Measure testing will focus on the following:

Validity: In order to determine the validity of each measure, aside from the face-validity in which the measures were built, validity testing will be conducted. Because the re-specified measures use the same clinical concepts, it is hypothesized they will perform similarly to the original EMS Compass candidate measures. The analysis for validity testing will include comparing aggregate agency/organizational-level measure scores of original EMS Compass candidate measures and the revised, re-specified measures.

Reliability: This effort will test the reliability of the overall measure performance. Reliability testing will determine if, as re-specified, the EMS Compass measures are able to produce consistent measure scores over time.

III. EMS Compass Measures to be tested

1. Hypoglycemia-01: Treatment Administered for Hypoglycemia
2. Pediatrics-01: Respiratory Assessment
3. Pediatrics-02: Administration of Beta Agonist for Pediatric Asthma
4. Pediatrics-03: Documentation of Estimated Weight in Kilograms
5. Seizure-02: Patient Received Intervention
6. Stroke-02: Suspected Stroke Receiving Prehospital Stroke Assessment
7. Trauma-01: Pain Assessment of Injured Patients
8. Trauma-03: Effectiveness of Pain Management for Injured Patients
9. Trauma-04: Trauma Patients Transported to the Trauma Center
10. Safety-01: Use of Lights and Sirens During Response to Scene
11. Safety-02: Use of Lights and Sirens During Transport

IV. Dataset: The datasets necessary for these analyses will initially require one data extract. If the workgroup decides to test a new measure element or modified specification during

the testing period, based on results of the first extract/analysis, additional extractions will be necessary. The data set and extract should be limited to the specified timeframe.

Percentile Report: Testing partners are asked to provide aggregate percentile reports for each measure, detailing 10th, 25th, 50th, and 90th percentiles using their 2018 data sets. Each percentile report should include aggregate data for all EMS encounters meeting inclusion criteria for each of the measures, or a random, statistically significant sample of encounters. In addition to the percentiles, testing partners are asked to include number of agencies/entities, number of EMS encounters, minimum, maximum, median, mean, and standard deviation on the data report.

Consistent with appropriate data use and privacy, extracts will not include Personal Health Information (PHI).

V. Data Management: The analysis dataset will be stored on a single, secure workstation belonging to the American College of Emergency Physicians. No attempt will be made to identify individual patients, providers, or agencies.

VI. Deliverables:

1. *Presentation*: A presentation of results will be prepared and delivered to the NEMSQA Measure Development Committee on 8/5/2019.

VII. Key Considerations:

1. Each testing partner will provide a complete dataset including aggregate data, as requested. No datasets will include identifying information.
2. Publication and Dissemination of overall results of testing will be at NEMSQA's discretion; however, individual data, measure scores, or calculations provided by testing partners will not be released.
3. This effort should be considered early testing and it is recommended that additional testing be completed after the re-specified measure set has been deployed for one year.

VIII. Timeline: This testing work will be conducted over a seven-week period

Week (date indicates start of each work week)	6/10	6/17	6/24	7/1	7/8	7/15	7/26	7/29
Finalize scope of work and testing plan								
Extract datasets								
Conduct initial analysis								
Implement potential specification changes (if necessary)								
Conduct additional analysis (if necessary)								
Prepare technical report and presentation								

IX. Project Team:

Testing & Learning Workgroup: includes both NEMSQA members and representatives from testing partners. Workgroup members include experts in the fields of EMS, quality, informatics, and measurement. The Testing & Learning workgroup will provide data, insights, and technical assistance throughout the project.

Michael Redlener, MD: is an EMS physician and NEMSQA Board of Trustee member who is overseeing the entire EMS Compass review and re-specification project.

*Kelly Burlison, MPH***:** will serve as the staff lead for this project, ensuring the project meets its milestones according to timeline, conducting analytics, and developing the final deliverables.

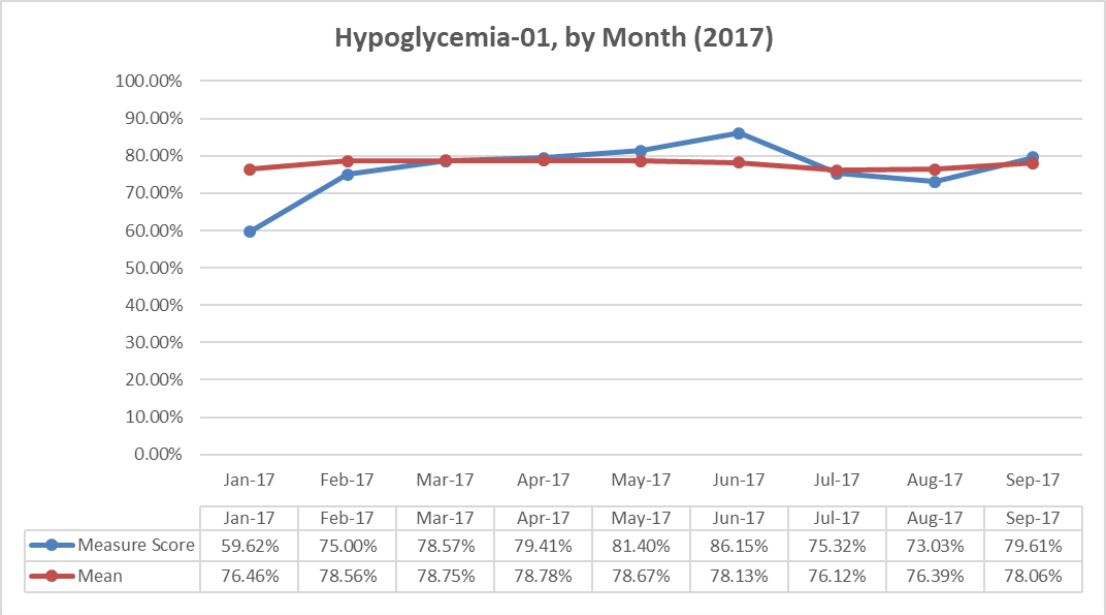
Testing Results

Below, data tables and charts are displayed, as available, that illustrate reliability and validity results for the re-specified EMS Compass Measures. While the measure testing plan was well laid out, limitations were faced during the testing period: such as the ability of some testing partners to provide data, the lack of certain data elements in the NEMSIS research dataset (considered a testing partner), and the strict, limited timeframe in which testing was conducted. However, the testing period did provide insightful results that were able to be tested. All measures that were able to be tested, proved to be reliable and valid.

Even though this round of testing produced positive results, NEMSQA intends to continue testing in 2020, after a full year of data have been collected.

Hypoglycemia-01: Reliability

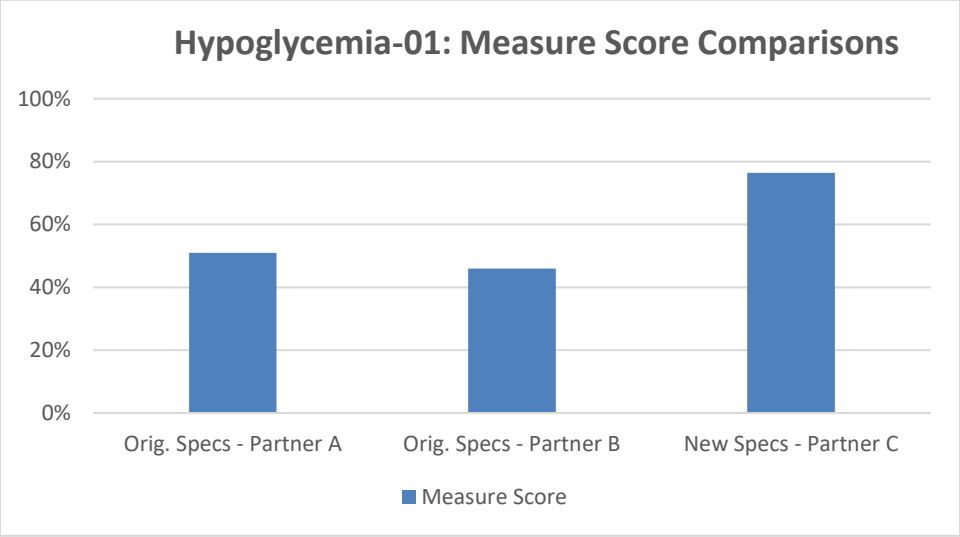
Hypoglycemia-01				
Month	**Numerator**	**Denominator**	**Measure Score**	**Z-Score**
Jan-17	31	52	59.62%	-2.07
Feb-17	30	40	75.00%	-0.06
Mar-17	33	42	78.57%	0.4
Apr-17	27	34	79.41%	0.51
May-17	35	43	81.40%	0.77
Jun-17	56	65	86.15%	1.4
Jul-17	58	77	75.32%	-0.02
Aug-17	65	89	73.03%	-0.32
Sep-17	82	103	79.61%	-0.54



	Jan-17	Feb-17	Mar-17	Apr-17	May-17	Jun-17	Jul-17	Aug-17	Sep-17
Measure Score	59.62%	75.00%	78.57%	79.41%	81.40%	86.15%	75.32%	73.03%	79.61%
Mean	76.46%	78.56%	78.75%	78.78%	78.67%	78.13%	76.12%	76.39%	78.06%

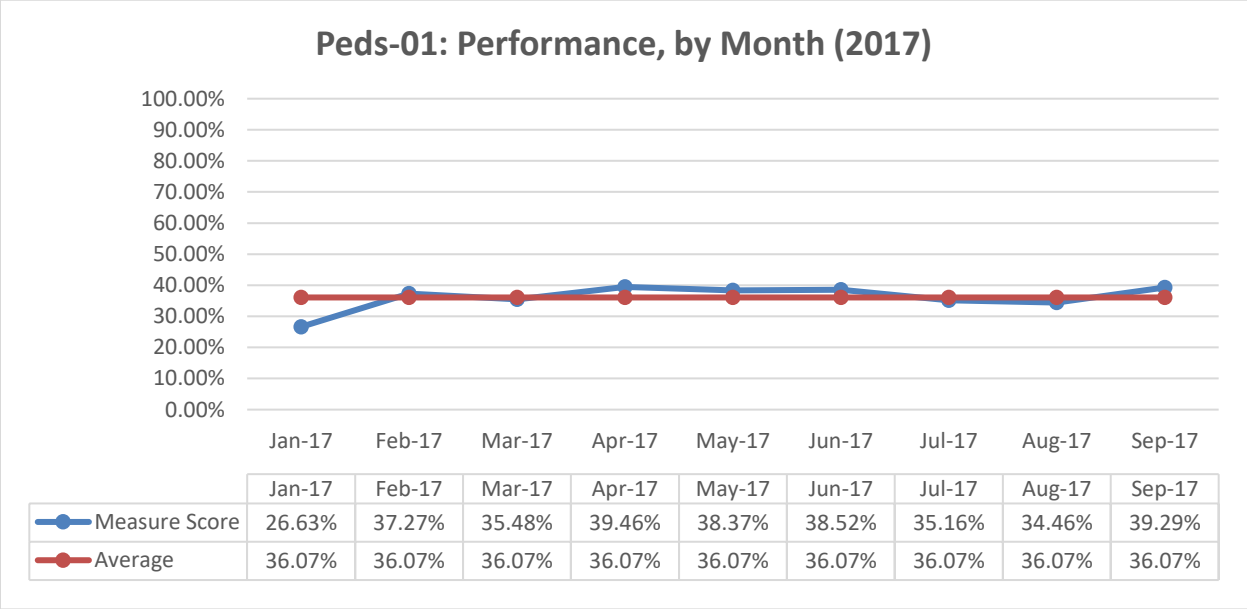
Hypoglycemia-01: Validity

Hypoglycemia-01: Measure Validity				
Sites/Specs	**Measure Score**	**Mean**	**Std Dev**	**Z-Score**
Orig. Specs - Partner A	51%	58%	0.163351523	-0.41
Orig. Specs - Partner B	46%	58%	0.163351523	-0.73
New Specs - Partner C	76%	58%	0.163351523	1.14



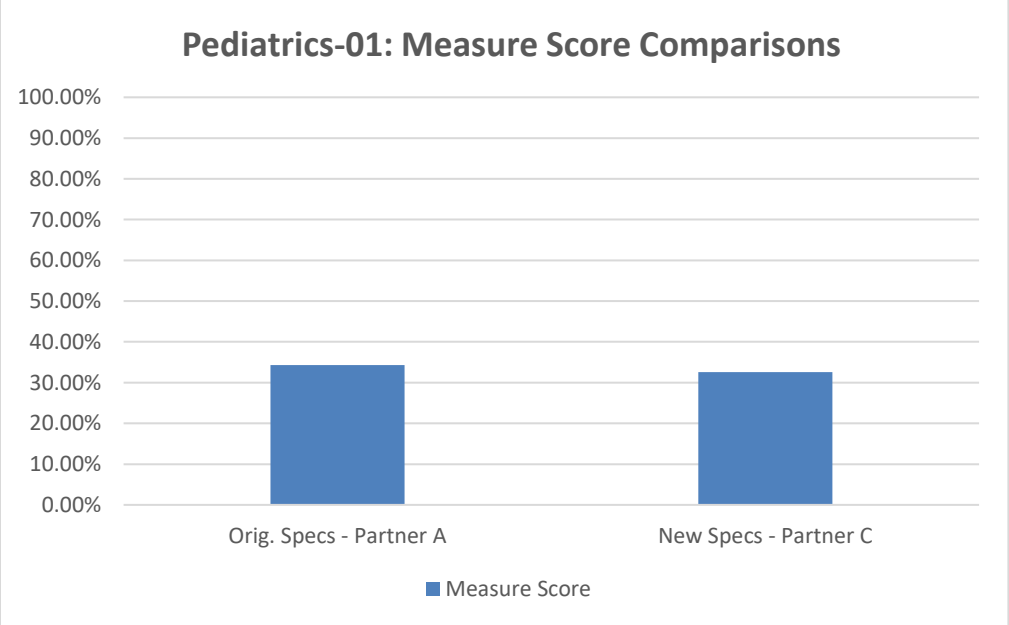
Pediatrics-01: Reliability

Pediatrics-01: Measure Reliability						
Month	**Numerator**	**Denominator**	**Measure Score**	**Mean**	**Standard Deviation**	**Z-Score**
Jan-17	94	353	26.63%	36.07%	0.039914242	-2.31
Feb-17	101	271	37.27%	36.07%	0.039914242	0.29
Mar-17	88	248	35.48%	36.07%	0.039914242	-0.23
Apr-17	88	223	39.46%	36.07%	0.039914242	0.81
May-17	94	245	38.37%	36.07%	0.039914242	0.55
Jun-17	52	135	38.52%	36.07%	0.039914242	0.81
Jul-17	45	128	35.16%	36.07%	0.039914242	-0.23
Aug-17	61	177	34.46%	36.07%	0.039914242	-0.49
Sep-17	77	196	39.29%	36.07%	0.039914242	0.81



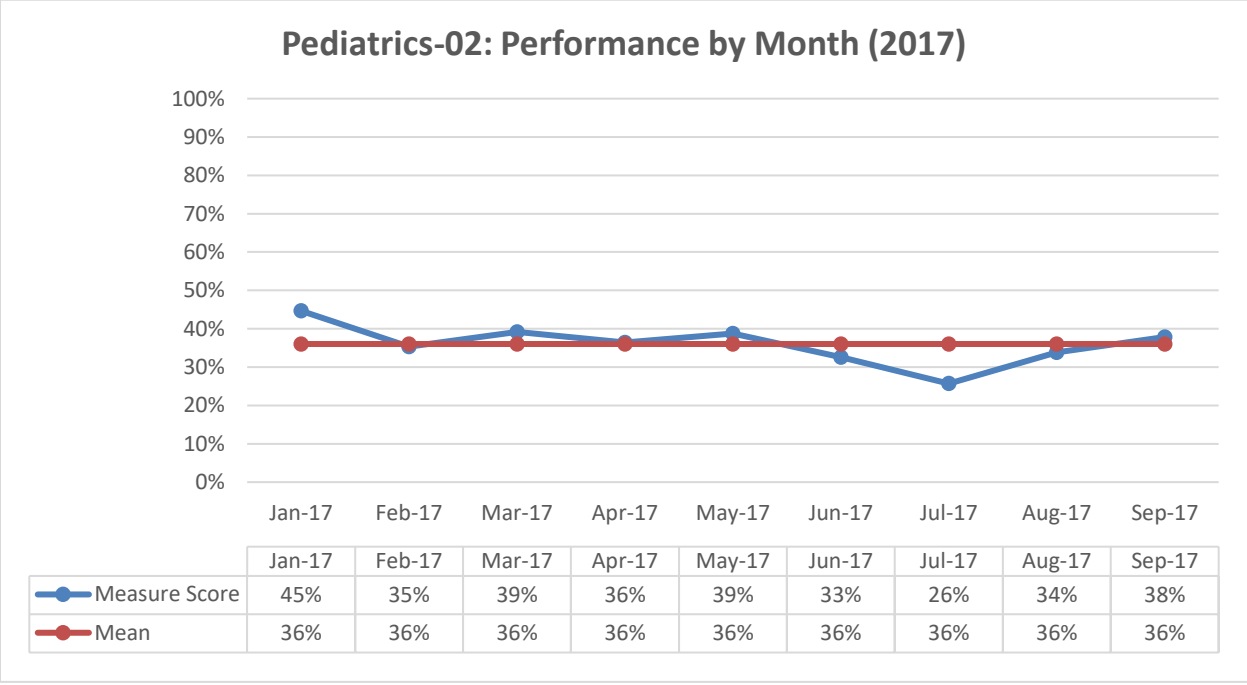
Pediatrics-01: Validity

Pediatricsc-01: Measure Validity				
Sites/Specs	**Measure Score**	**Mean**	**Std Dev**	**Z-Score**
Orig. Specs - Partner A	34.31%	57.82%	0.012020815	0.71
New Specs - Partner C	32.61%	57.82%	0.012020815	-0.71



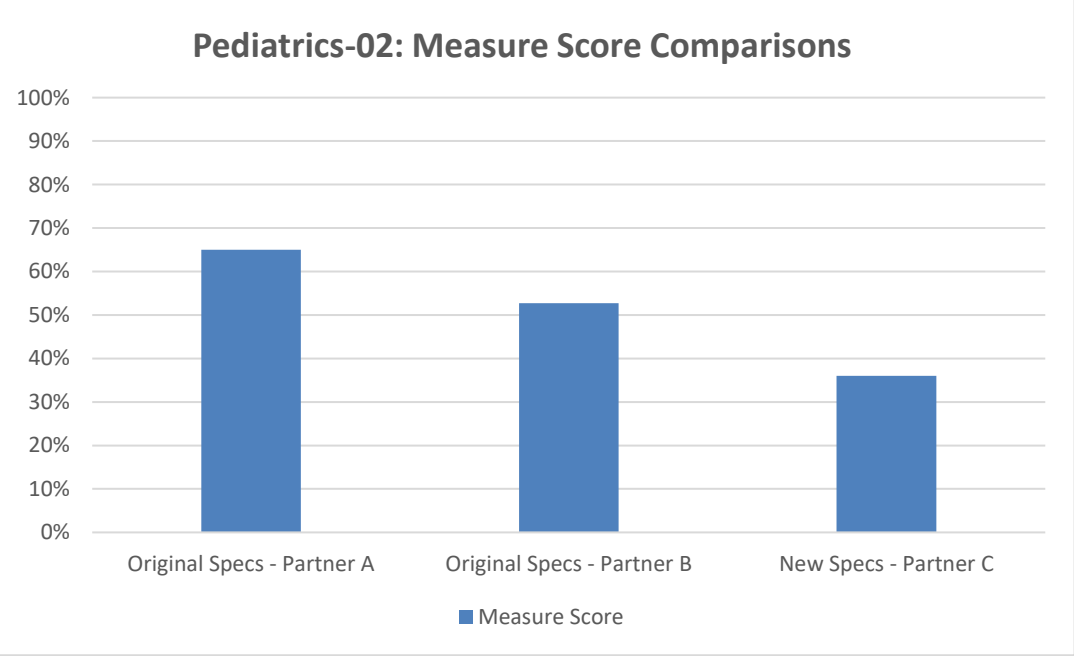
Pediatrics-02: Reliability

Pediatrics-02 - Measure Reliability						
Month	Numerator	Denominator	Measure Score	Mean	Standard Deviation	Z-Score
Jan-17	138	309	45%	36%	0.052306568	1.65
Feb-17	94	266	35%	36%	0.052306568	-0.13
Mar-17	99	253	39%	36%	0.052306568	0.6
Apr-17	90	247	36%	36%	0.052306568	0.08
May-17	98	253	39%	36%	0.052306568	0.52
Jun-17	55	169	33%	36%	0.052306568	-0.66
Jul-17	37	144	26%	36%	0.052306568	1.97
Aug-17	65	192	34%	36%	0.052306568	-0.41
Sep-17	96	254	38%	36%	0.052306568	0.34



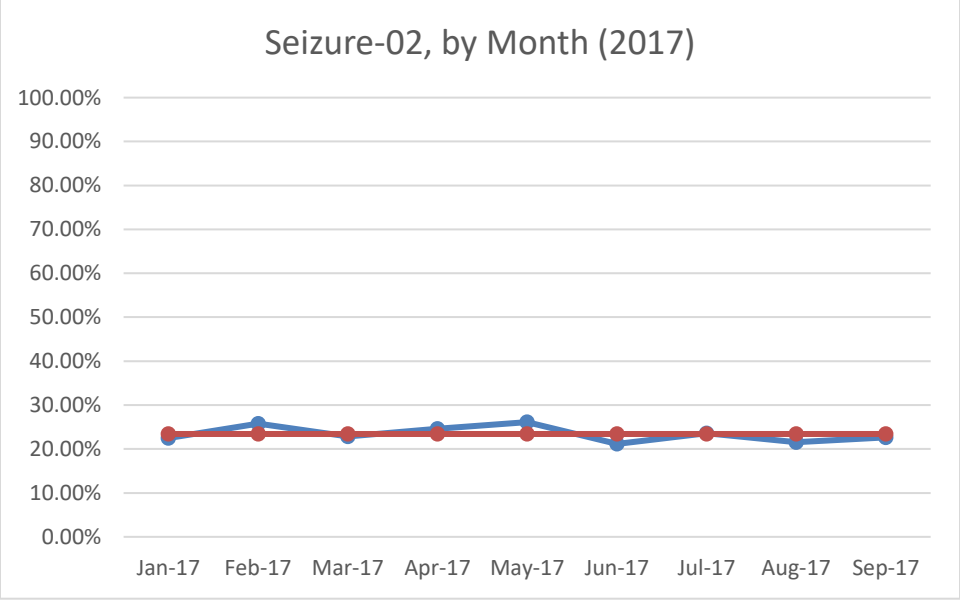
Pediatrics-02: Validity

Pediatrics-02: Measure Validity				
Specs/Site	**Measure Score**	**Mean**	**Std Dev**	**Z-Score**
Original Specs - Partner A	65%	51%	0.14545061	0.94
Original Specs - Partner B	53%	51%	0.14545061	0.11
New Specs - Partner C	36%	51%	0.14545061	-1.05



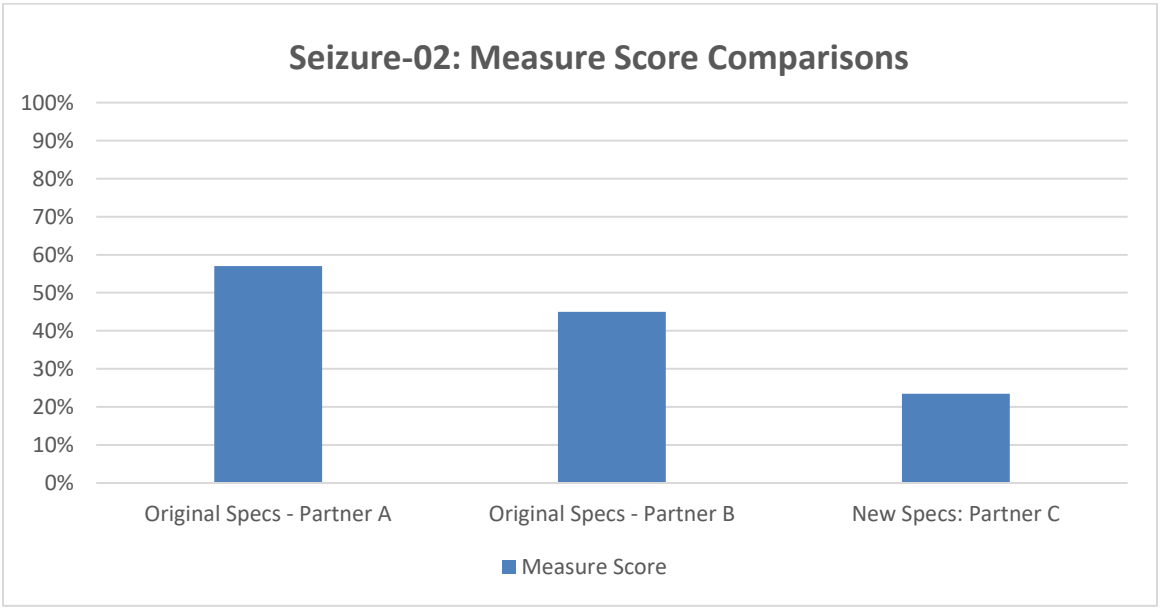
Seizure-02: Reliability

Seizure-02 - Measure Reliability						
Month	**Numerator**	**Denominator**	**Measure Score**	**Mean**	**Standard Deviation**	**Z-Score**
Jan-17	40	178	22.47%	23.42%	0.017565053	-0.54
Feb-17	34	132	25.76%	23.42%	0.017565053	1.33
Mar-17	38	166	22.89%	23.42%	0.017565053	-0.3
Apr-17	33	134	24.63%	23.42%	0.017565053	0.59
May-17	47	180	26.11%	23.42%	0.017565053	1.53
Jun-17	33	156	21.15%	23.42%	0.017565053	-1.29
Jul-17	45	191	23.56%	23.42%	0.017565053	0.08
Aug-17	31	144	21.53%	23.42%	0.017565053	-1.07
Sep-17	36	159	22.64%	23.42%	0.017565053	-0.44



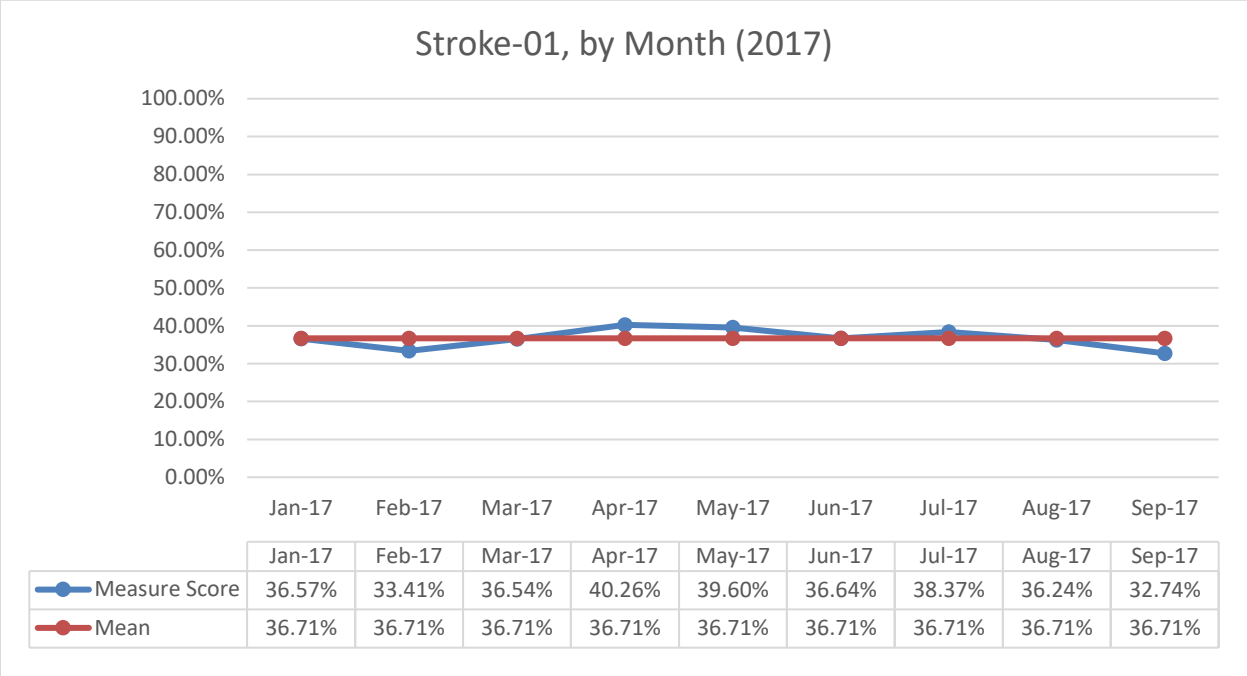
Seizure-02: Validity

Seizure-02: Measure Validity				
Site/Specs	**Measure Score**	**Mean**	**Std Dev**	**Z-Score**
Original Specs - Partner A	57%	42%	0.170162315	0.89
Original Specs - Partner B	45%	42%	0.170162315	0.19
New Specs: Partner C	23%	42%	0.170162315	-1.08



Stroke-01: Reliability

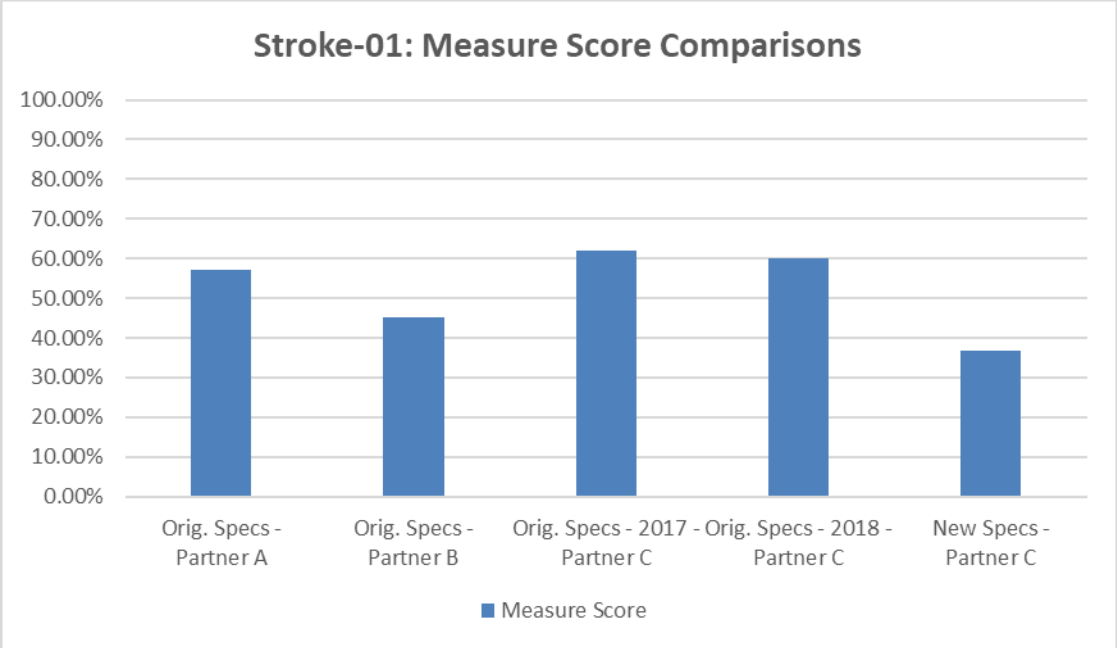
Stroke-01 - Measure Reliability						
Month	**Numerator**	**Denominator**	**Measure Score**	**Mean**	**Standard Deviation**	**Z-Score**
Jan-17	196	536	36.57%	36.71%	0.025166676	0.09
Feb-17	152	455	33.41%	36.71%	0.025166676	-1.49
Mar-17	171	468	36.54%	36.71%	0.025166676	0.09
Apr-17	184	457	40.26%	36.71%	0.025166676	1.27
May-17	177	447	39.60%	36.71%	0.025166676	1.27
Jun-17	155	423	36.64%	36.71%	0.025166676	0.09
Jul-17	155	404	38.37%	36.71%	0.025166676	0.48
Aug-17	154	425	36.24%	36.71%	0.025166676	-0.31
Sep-17	110	336	32.74%	36.71%	0.025166676	-1.49



	Jan-17	Feb-17	Mar-17	Apr-17	May-17	Jun-17	Jul-17	Aug-17	Sep-17
Measure Score	36.57%	33.41%	36.54%	40.26%	39.60%	36.64%	38.37%	36.24%	32.74%
Mean	36.71%	36.71%	36.71%	36.71%	36.71%	36.71%	36.71%	36.71%	36.71%

Stroke-01: Validity

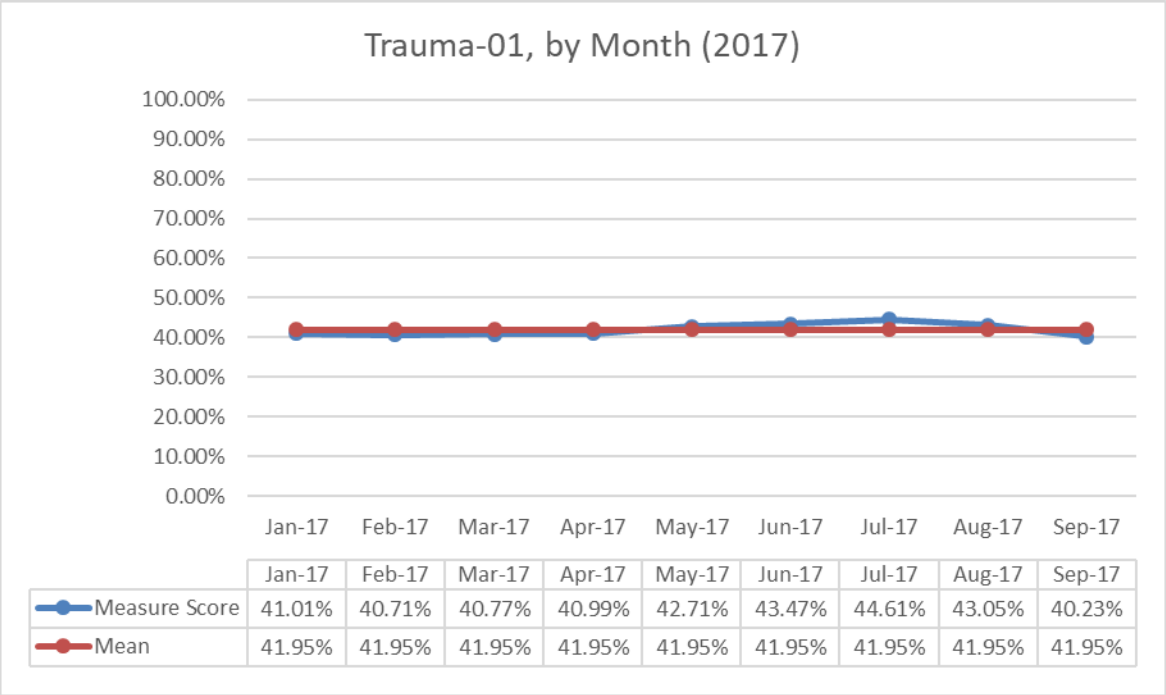
Stroke-01 - Measure Validity				
Specs/Site	**Measure Score**	**Mean**	**Standard Deviation**	**Z-Score**
Orig. Specs - Partner A	57.00%	52.23%	0.107675847	0.44
Orig. Specs - Partner B	45.30%	52.23%	0.107675847	-0.65
Orig. Specs - 2017 - Partner C	62%	52.23%	0.107675847	0.91
Orig. Specs - 2018 - Partner C	60%	52.23%	0.107675847	0.72
New Specs - Partner C	36.83%	52.23%	0.107675847	-1.43



Trauma-01: Reliability

Trauma-01 - Measure Reliability						
Month	**Numerator**	**Denominator**	**Measure Score**	**Mean**	**Standard Deviation**	**Z-Score**
Jan-17	3189	7776	41.01%	41.95%	0.015368451	-0.63
Feb-17	2562	6294	40.71%	41.95%	0.015368451	-0.63
Mar-17	2816	6907	40.77%	41.95%	0.015368451	-0.63
Apr-17	2904	7085	40.99%	41.95%	0.015368451	-0.63
May-17	3270	7656	42.71%	41.95%	0.015368451	0.63

Jun-17	3145	7235	43.47%	41.95%	0.015368451	0.63
Jul-17	3473	7785	44.61%	41.95%	0.015368451	1.9
Aug-17	3064	7117	43.05%	41.95%	0.015368451	0.63
Sep-17	2458	6110	40.23%	41.95%	0.015368451	-1.26

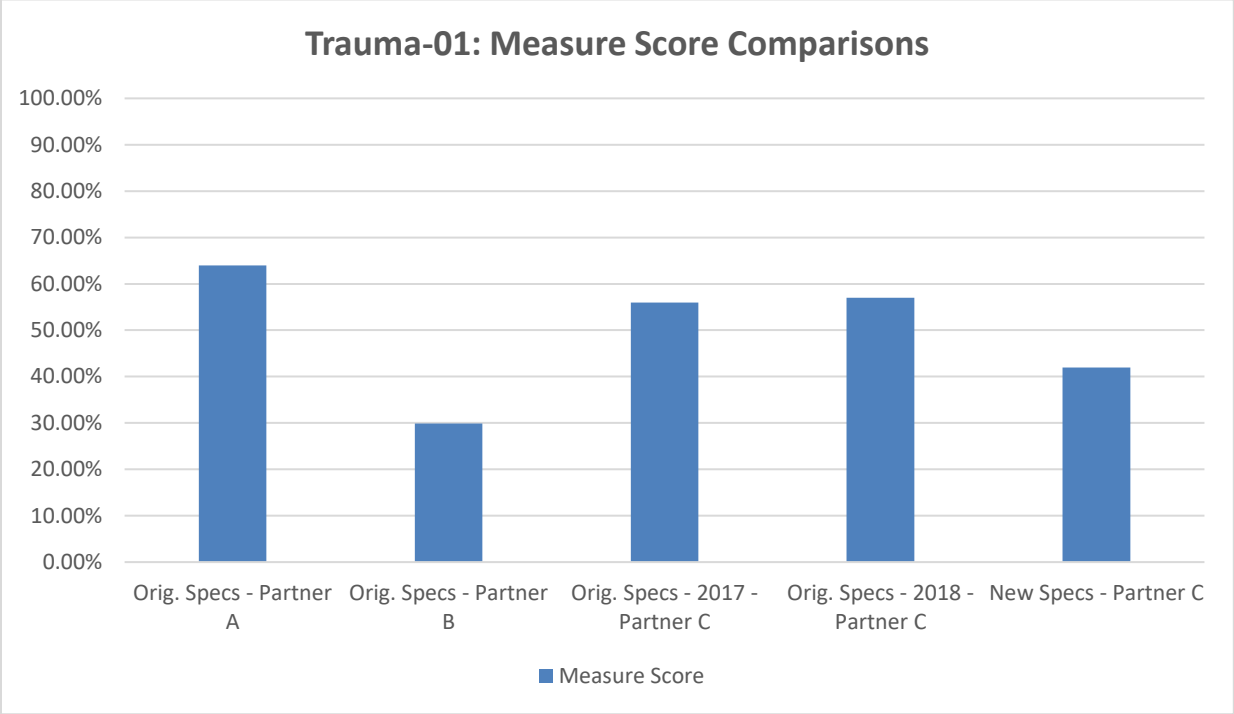


	Jan-17	Feb-17	Mar-17	Apr-17	May-17	Jun-17	Jul-17	Aug-17	Sep-17
Measure Score	41.01%	40.71%	40.77%	40.99%	42.71%	43.47%	44.61%	43.05%	40.23%
Mean	41.95%	41.95%	41.95%	41.95%	41.95%	41.95%	41.95%	41.95%	41.95%

Trauma-01: Validity

Trauma-01 - Measure Validity				
Specs/Site	**Measure Score**	**Mean**	**Standard Deviation**	**Z-Score**
Orig. Specs - Partner A	64.00%	49.77%	0.136889371	1.04
Orig. Specs - Partner B	29.90%	49.77%	0.136889371	-1.45
Orig. Specs - 2017 - Partner C	56.00%	49.77%	0.136889371	0.45

Orig. Specs - 2018 - Partner C	57.00%	49.77%	0.136889371	0.53
New Specs - Partner C	41.95%	49.77%	0.136889371	-0.57



Trauma-04: Reliability

Trauma-04 - Measure Reliability					
Month	**Numerator**	**Denominator**	**Measure Score**	**Mean**	**Z-Score**
Jan-17	203	1408	14.42%	14.96%	-0.25

Feb-17	195	1405	13.88%	14.96%	-0.25
Mar-17	191	1599	11.94%	14.96%	-0.75
Apr-17	216	1699	12.71%	14.96%	-0.50
May-17	159	728	21.84%	14.96%	1.75