

Analysis of the  
University of Auckland MBChB  
Progress Test 29 Data  
(Report for Students)

Nathan Ryckman  
Nikki Tugnet  
Medical Programme Directorate  
University of Auckland

July 26, 2022

## Contents

<b>1</b>	<b>Test Information</b>	<b>2</b>
<b>2</b>	<b>Student Progress</b>	<b>2</b>
<b>3</b>	<b>Test Statistics</b>	<b>4</b>
3.1	Summary Statistics . . . . .	4
3.2	Grade Boundaries . . . . .	7
3.3	Response Summary . . . . .	8
3.3.1	PT29 Data . . . . .	8
3.3.2	Longitudinal Data . . . . .	10

## List of Tables

1	Current Year 2 Descriptive Statistics . . . . .	4
2	Current Year 3 Descriptive Statistics . . . . .	4
3	Current Year 4 Descriptive Statistics . . . . .	4
4	Current Year 5 Descriptive Statistics . . . . .	4
5	Current Year 6 Descriptive Statistics . . . . .	5
6	Distribution of students per cohort that will be awarded different PT grades . . . . .	7
7	Cut Scores . . . . .	7
8	Mean Number of Responses to 120 Items . . . . .	8
9	Percentage of Responses . . . . .	10

## List of Figures

1	Longitudinal Progress Test Scores (Mean and 95% CI) . . . . .	3
2	Boxplots of PT29 Scores By Cohort . . . . .	6
3	Response Summary Boxplots . . . . .	8
4	Response Summary Boxplots by Cohort and PT29 Grade . . . . .	9
5	Summaries of Responses . . . . .	11

## 1 Test Information

This document was prepared for students who sat Progress Test 29. It contains the results of the **FINAL** analysis run of the University of Auckland MBChB Progress Test 29 (PT29) data. The final analysis run excludes any questions that were deemed problematic in the post test moderation meeting. The following items were excluded in this run : Q020, Q026, Q022, Q076, Q086

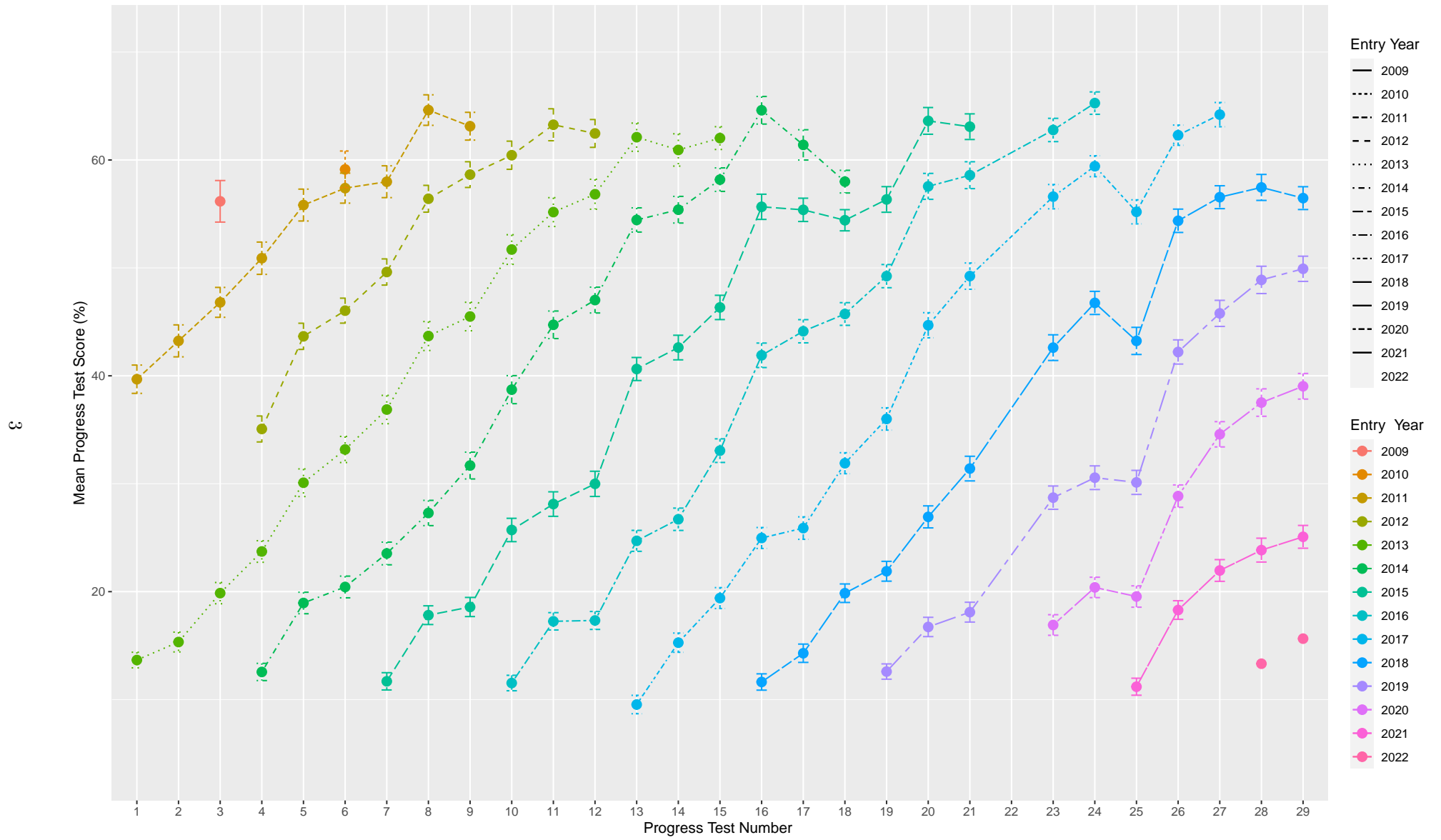
The test date was 15/07/2022.

## 2 Student Progress

Progress Testing began in 2013 with PT1. Since then, three progress tests have been conducted each year, with the exception of 2020. In 2020, the first test of the year (PT22) was cancelled as a result of the restrictions imposed in response to COVID-19. Figure 1.<sup>1</sup> on the following page shows the mean progress test score for each year group in each test. It should be noted that the mean scores reflect the combination of student ability and test difficulty. While a year group's mean test score should increase over time if all tests are equally difficult, harder tests may lead to some or all year groups showing lower mean scores in specific tests. Variation in test difficulty is accounted for through setting specific grade boundaries for each test, as outlined in section 3.2. It is also worth noting that the mean scores are based on the calculated score, where incorrect responses incur a -0.25 mark penalty.

---

<sup>1</sup>Year 6 students in 2013 and 2014 also sat PT03 (n = 123) and PT06 (n = 138) as formative tests, respectively. The data from these groups (Cohorts 2009 and 2010) inform the standards that are set for Y6 students, starting with PT07.



Date Created: 2022-07-26 15:49:09

Figure 1: Longitudinal Progress Test Scores (Mean and 95% CI)

### 3 Test Statistics

#### 3.1 Summary Statistics

Table 1 to Table 5 show the summary statistics of the scores in all progress tests that the current Year 2 to Year 6 students have sat.

Table 1: Current Year 2 Descriptive Statistics

PT	Minimum	P25	Median	Mean	SD	P75	P95	Maximum	IQR	Range	N
PT28	-0.4	7.8	11.7	13.3	7.6	17.2	28.9	41.6	9.4	42.0	283.0
PT29	-0.6	10.2	14.2	15.6	7.7	19.8	31.5	42.5	9.6	43.1	276.0

Table 2: Current Year 3 Descriptive Statistics

PT	Minimum	P25	Median	Mean	SD	P75	P95	Maximum	IQR	Range	N
PT25	-3.5	7.1	10.2	11.2	6.8	14.2	25.1	40.2	7.1	43.7	293.0
PT26	0.6	12.9	17.4	18.3	7.5	21.7	32.2	45.7	8.8	45.1	291.0
PT27	3.3	16.0	21.9	22.0	8.7	27.3	35.1	58.8	11.3	55.5	292.0
PT28	6.4	17.4	23.0	23.8	9.4	29.5	40.9	61.5	12.1	55.1	283.0
PT29	4.8	19.4	24.2	25.1	9.0	30.2	41.6	59.0	10.8	54.2	283.0

Table 3: Current Year 4 Descriptive Statistics

PT	Minimum	P25	Median	Mean	SD	P75	P95	Maximum	IQR	Range	N
PT23	-0.4	10.9	16.4	16.9	8.1	21.9	31.3	45.9	11.0	46.3	282.0
PT24	2.2	14.6	19.7	20.4	8.1	25.7	34.2	52.4	11.1	50.2	282.0
PT25	1.5	13.7	18.8	19.5	8.4	25.3	34.4	49.8	11.6	48.3	277.0
PT26	9.6	22.1	28.1	28.9	8.7	34.8	44.8	58.2	12.7	48.6	274.0
PT27	12.1	27.3	33.5	34.6	9.8	41.0	53.6	62.1	13.7	50.0	274.0
PT28	13.5	29.3	36.3	37.5	10.8	44.7	56.6	66.2	15.4	52.7	279.0
PT29	6.7	32.1	38.3	39.0	10.0	45.6	57.0	67.9	13.5	61.2	274.0

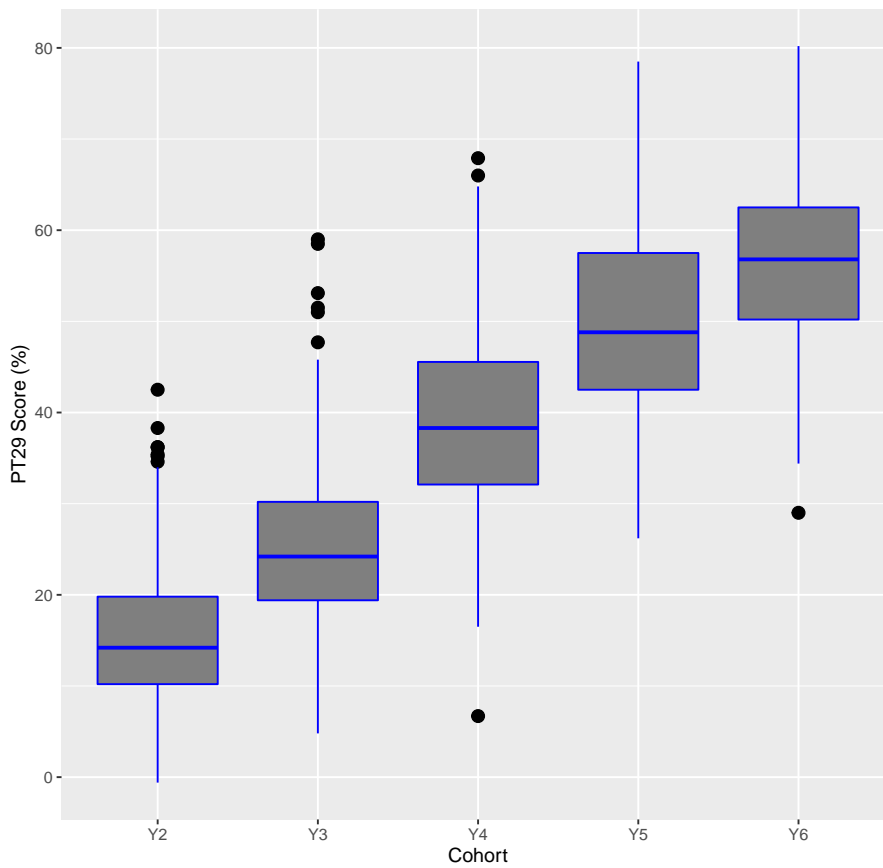
Table 4: Current Year 5 Descriptive Statistics

PT	Minimum	P25	Median	Mean	SD	P75	P95	Maximum	IQR	Range	N
PT19	-0.4	8.7	12.0	12.6	6.0	15.7	22.6	38.2	7.0	38.6	276.0
PT20	-3.9	11.5	15.6	16.7	7.6	20.9	27.9	57.6	9.4	61.5	278.0
PT21	1.3	12.9	17.4	18.1	7.7	22.2	33.4	47.9	9.3	46.6	275.0
PT23	8.8	22.3	27.7	28.7	9.1	33.8	44.7	69.7	11.5	60.9	276.0
PT24	10.2	24.1	29.8	30.6	9.2	35.8	47.4	65.9	11.7	55.7	274.0
PT25	9.0	23.7	29.8	30.1	9.4	36.0	44.2	68.1	12.3	59.1	274.0
PT26	17.8	36.2	42.2	42.2	9.5	48.2	58.4	77.7	12.0	59.9	276.0
PT27	21.1	38.3	45.3	45.8	10.2	52.9	64.5	73.8	14.6	52.7	277.0
PT28	9.6	41.0	48.8	48.9	10.6	56.4	65.7	77.7	15.4	68.1	270.0
PT29	26.2	42.5	48.8	49.9	9.7	57.7	67.0	78.5	15.2	52.3	267.0

Table 5: Current Year 6 Descriptive Statistics

PT	Minimum	P25	Median	Mean	SD	P75	P95	Maximum	IQR	Range	N
PT16	-0.4	6.9	10.8	11.6	6.5	15.6	23.1	39.8	8.7	40.2	284.0
PT17	1.4	8.8	13.5	14.3	7.2	18.7	27.3	41.6	9.9	40.2	282.0
PT18	3.1	14.9	19.4	19.9	7.3	24.4	32.4	47.1	9.5	44.0	281.0
PT19	2.2	16.9	20.9	21.9	7.8	26.3	36.0	50.6	9.4	48.4	280.0
PT20	10.9	21.1	26.4	26.9	8.7	31.6	42.9	54.3	10.5	43.4	280.0
PT21	7.6	23.8	30.9	31.4	9.7	37.9	48.9	64.0	14.1	56.4	278.0
PT23	19.1	34.8	42.0	42.6	10.1	50.0	60.9	68.2	15.2	49.1	279.0
PT24	25.8	40.4	46.1	46.8	9.1	52.9	62.0	73.6	12.5	47.8	278.0
PT25	16.7	35.6	42.3	43.2	10.5	50.6	62.2	75.0	15.0	58.3	271.0
PT26	25.8	48.0	54.0	54.4	9.1	61.7	70.2	78.5	13.7	52.7	272.0
PT27	35.5	50.4	55.9	56.5	9.1	63.1	70.9	82.8	12.7	47.3	282.0
PT28	14.5	50.6	57.0	57.5	10.1	64.8	73.0	81.6	14.2	67.1	274.0
PT29	29.0	50.2	56.8	56.5	8.8	62.5	71.7	80.2	12.3	51.2	270.0

The diagram in Figure 2 visually captures some of the key summary statistics for PT29 scores that are found in the descriptive statistics tables. The horizontal line inside the box represents the median PT29 score. The top and bottom borders of the box show the 75th percentile (P75) and 25th percentile (P25), respectively. The height of the box shows the the interquartile range (IQR). Outliers are defined as any value more than 1.5 times the interquartile range from the 25th/75th percentile (whichever is closer) and are shown as black dots. The ends of the whiskers correspond to the minimum and maximum PT29 scores, excluding outliers. The lowest point to the highest point within a year group (including the outliers) shows the range of the scores in the distribution.



Date Created: 2022-07-26 15:49:09

Figure 2: Boxplots of PT29 Scores By Cohort

### 3.2 Grade Boundaries

Because Year 2 to Year 5 students are graded normatively, certain percentages of the cohort are expected to be classified in one of the four PT grades as shown in Table 6. The cut-scores that would yield PT29 grades with this distribution are found in Table 7.

Table 6: Distribution of students per cohort that will be awarded different PT grades

Cohort	Unsatisfactory	Borderline	Satisfactory	Excellent
Year 2	5%	10%	80%	5%
Year 3	4%	8%	83%	5%
Year 4	3%	6%	86%	5%
Year 5	2%	4%	89%	5%

Table 7: Cut Scores

Cohort	Boundary	Cut Score
Y2	UB	4.8
Y2	BS	8.5
Y2	SE	31.5
Y3	UB	10.4
Y3	BS	15.2
Y3	SE	41.4
Y4	UB	22.6
Y4	BS	26.9
Y4	SE	57.1
Y5	UB	33.1
Y5	BS	36.2
Y5	SE	67.0

Year 6 boundary scores were determined using the item-mapping method of standard setting as proposed by Wang (2003)<sup>2</sup>. The results were triangulated using historical data as well as Cohen’s method (Cohen-Schotanus and van der Vleuten, 2010). Number-correct scores, instead of formula (or negatively marked) scores were analysed using Rasch Analysis software to estimate the locations of the items (*difficulty*) and of test takers (*ability*) on the same scale.

The two groups of experts that participated in the item-mapping sessions were composed of clinicians familiar with the standard expected of both final year medical students and new graduates. The results of both standard setting sessions were considered to propose that the PT29 score of 39.2% (when using negative marking for incorrect responses) be the Pass-Fail cut-score.

The item-mapping method is a recognised method for determining the Pass-Fail cut-score. However, experts may find it difficult to find a consensus when determining the Excellent-Pass cut-score, and different groups of experts are more likely to propose very different cut-scores for Excellent-Pass boundaries than pass/fail boundaries. As a result, the Excellent-Pass cut-score, with the approval of the Assessment Subcommittee, was determined using norm-referencing. Just like in other cohorts, the progress test grade that separated the top 5% of the cohort became the cut-score. With the Year 6 cohort, the PT29 cut-score was 71.7%.

<sup>2</sup>See MacCann and Stanley (2006) for comparison of item mapping with the more familiar methods such as Angoff procedure and the bookmark method.

### 3.3 Response Summary

#### 3.3.1 PT29 Data

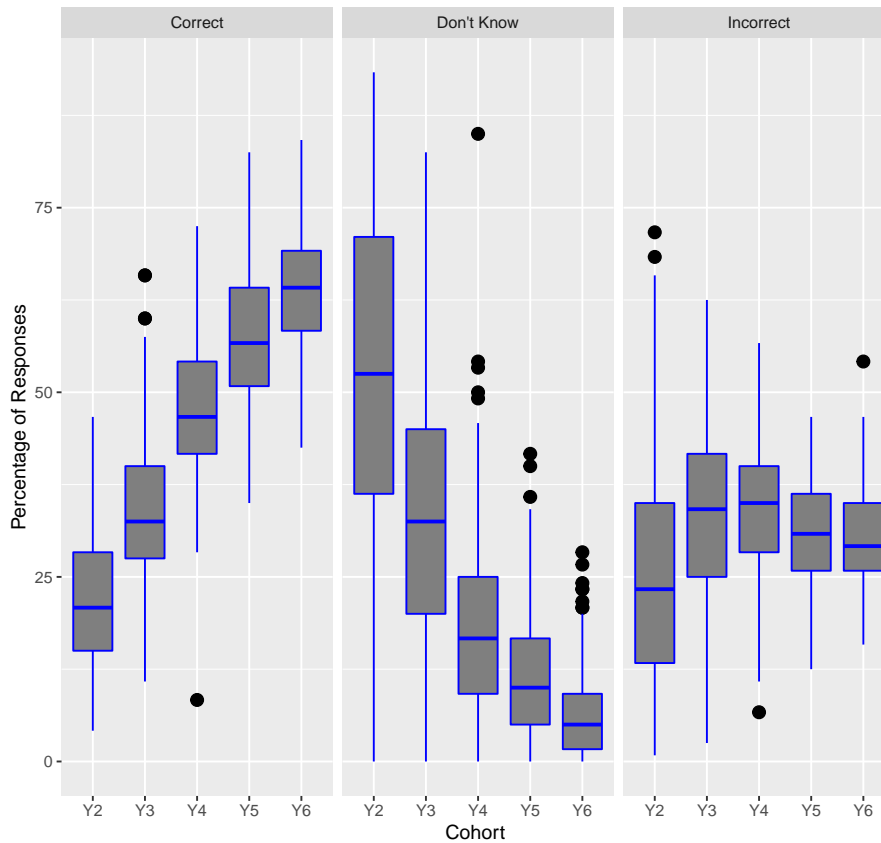
Table 8 shows the average counts of the three possible response types to the 120 PT29 items: correct, incorrect and *don't know* (DK) responses.

In progress testing, students who are relatively new to the programme are, on average, expected to have more DK and fewer correct responses than those who have been in the programme for a longer period of time.

Table 8: Mean Number of Responses to 120 Items

Response	Y2	Y3	Y4	Y5	Y6
Correct	26.6	40.2	57.1	69.1	76.7
Don't Know	62.2	39.4	21.8	14.0	7.4
Incorrect	31.2	40.4	41.1	36.9	35.9

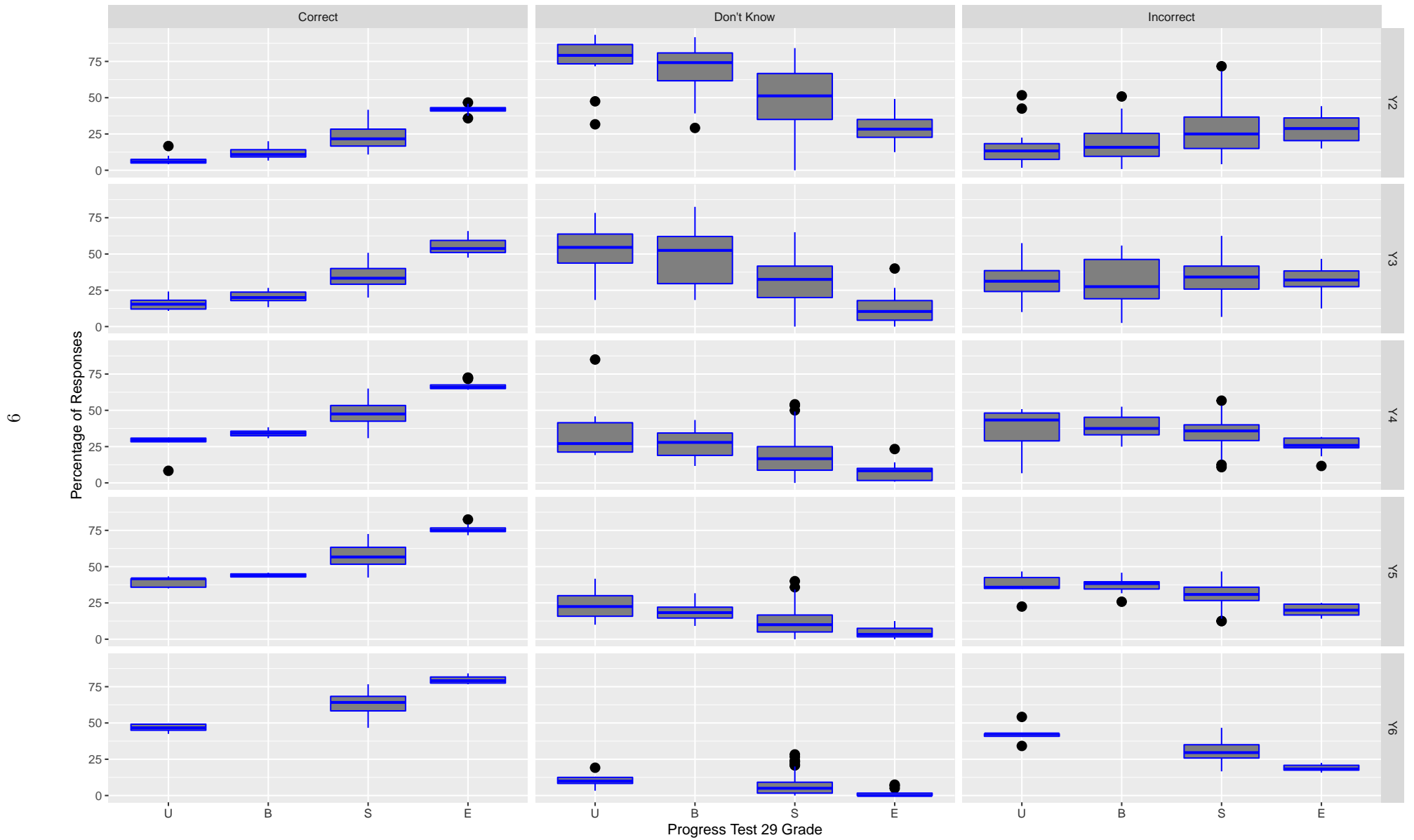
Figure 3 shows the summary statistics *minimum*, *P25*, *median*, *P75* and *maximum* for each response type per cohort. Boxplots that show the same summary statistics for each response type per PT29 grade of each cohort are found in Figure 4.



Date Created: 2022-07-26 15:49:09

Figure 3: Response Summary Boxplots





Date Created: 2022-07-26 15:49:09

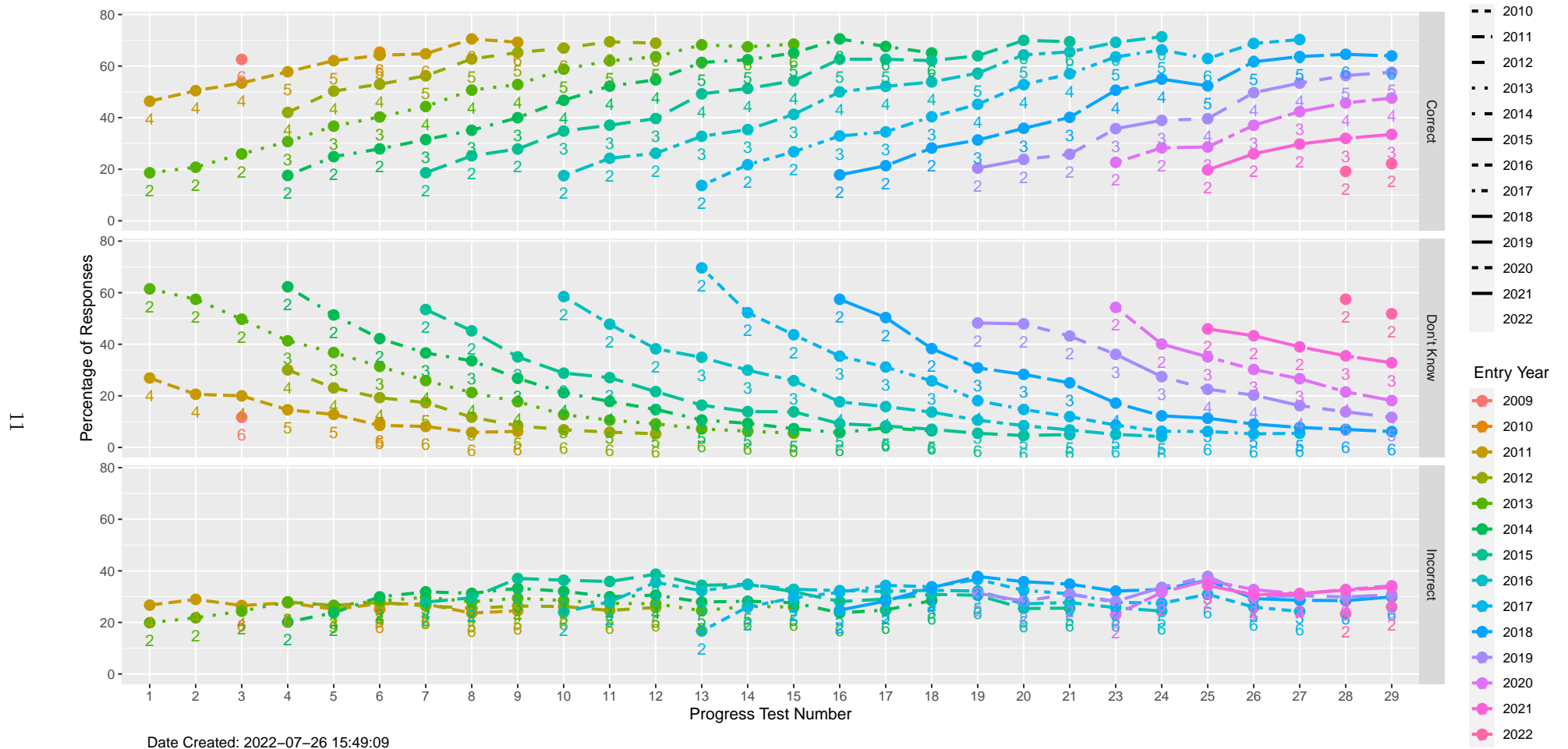
Figure 4: Response Summary Boxplots by Cohort and PT29 Grade

### 3.3.2 Longitudinal Data

Table 9: Percentage of Responses

Cohort	Response	PT15	PT16	PT17	PT18	PT19	PT20	PT21	PT23	PT24	PT25	PT26	PT27	PT28	PT29
2022	Correct													19	22
2022	Don't Know													57	52
2022	Incorrect													23	26
2021	Correct										20	26	30	32	33
2021	Don't Know										46	43	39	36	33
2021	Incorrect										34	31	31	33	34
2020	Correct								23	28	29	37	42	46	48
2020	Don't Know								54	40	35	30	27	22	18
2020	Incorrect								23	32	36	33	31	33	34
2019	Correct					20	24	26	36	39	40	50	53	56	58
2019	Don't Know					48	48	43	36	27	23	20	16	14	12
2019	Incorrect					31	28	31	28	34	38	30	30	30	31
2018	Correct		18	21	28	31	36	40	51	55	52	62	64	65	64
2018	Don't Know		57	50	38	31	28	25	17	12	11	9	8	7	6
2018	Incorrect		25	28	33	38	36	35	32	33	36	29	29	28	30
2017	Correct	27	33	34	40	45	53	57	64	66	63	69	70		
2017	Don't Know	44	35	31	26	18	15	12	9	6	6	5	5		
2017	Incorrect	30	32	34	34	37	32	31	28	27	31	26	24		
2016	Correct	41	50	52	54	57	64	66	69	71					
2016	Don't Know	26	18	16	14	11	8	7	5	4					
2016	Incorrect	33	32	32	32	32	27	28	26	24					
2015	Correct	54	63	63	62	64	70	69							
2015	Don't Know	14	9	8	7	5	5	5							
2015	Incorrect	32	28	29	31	31	25	26							
2014	Correct	65	71	68	65										
2014	Don't Know	7	6	8	6										
2014	Incorrect	28	24	25	28										
2013	Correct	69													
2013	Don't Know	6													
2013	Incorrect	26													

Figure 5 shows the summaries of correct, don't know, and incorrect responses from the first to the current progress test.



Date Created: 2022-07-26 15:49:09

Figure 5: Summaries of Responses

## References

- Cohen-Schotanus, J. and van der Vleuten, C. P. M. (2010). A standard setting method with the best performing students as point of reference: Practical and affordable. *Medical Teacher*, 32(2):154–160. <http://dx.doi.org/10.3109/01421590903196979>.
- MacCann, R. G. and Stanley, G. (2006). The use of Rasch modeling to improve standard setting. *Practical Assessment Research & Evaluation*, 11(2):1–17. Available online: <http://pareonline.net/getvn.asp?v=11&n=2>.
- Wang, N. (2003). Use of the Rasch IRT model in standard setting: An item-mapping method. *Journal of Educational Measurement*, 40(3):231–253. <https://goo.gl/1J0aua>.