

# IB SCORE REPORTS

## Case Study #2

How Our Display of a Moderation Trend Led to Higher Scores

In this case study, we show how our displays helped a school discover an unsettling five-year trend in the moderation of their Chemistry HL internal assessment scores. Subsequently, both scoring accuracy and student scores improved markedly.

*This case study is based on reports from IB Score Reports.com.*

*Although the data presented are real, the name of the school and the subject have been changed.*

### 1. Five Years of Consistent Downward Moderation

We begin by looking at the Moderation Summary Table for Chemistry HL at Acadamigo International School (next page). The table presents an overall picture of how internal assessment scores for that subject at that school have been moderated by the IBO over the five year period, 2010 to 2014.

The scores running down the left of the table represent the range of raw scores submitted by the school from 2010 to 2014. The highest value, 48, is the highest raw score the school submitted, and the lowest value, 16, is the lowest raw score the school submitted.

In the column for each year, we can see how the IBO moderated the raw scores submitted by the school. Focusing on the upper rows of the table, one can see that **the school's raw scores have been substantially downwardly moderated year after year.**

For example, in 2010 raw scores of 48 were moderated down 12 points and raw scores of 46 were moderated down 11 points. The moderation in 2013 was more extreme. Raw scores of 48 were moderated down 18 points, and raw scores of 47 and 46 were moderated down 17 points. In 2014, the school did not submit any scores of 48 or 47, which is why those cells are blank. The highest score the school submitted in 2014 was a 46, and that was moderated down 15 points.

**It is important to remember that moderation applies to all students earning a given score.**

For example, in 2010 every student at Acadamigo International School who was awarded a score of 48 points on the Chemistry HL internal assessment was moderated down 12 points by the IBO.

Returning to the full display, one can see that nearly every raw score submitted by the school for this assessment over the five-year period was moderated down by the IBO, and that none were moderated up.

CHEMISTRY HL PRACTICAL WORK

\*How to Read This Table

Raw	2010	2011	2012	2013	2014
48	-12			-18	
47			-11	-17	
46	-11			-17	-15
45			-10		
44					-14
43				-15	-13
42	-10	-6			
41			-9		
40				-13	-11
39		-5			
38	-8			-11	
37				-11	
36				-10	-8
35			-7	-9	-7
34					
33	-6	-4	-6		-5
32		-4			-5
31				-7	-4
30	-5	-4		-6	
29				-6	
28	-5				
27	-4				
26	-4				
25		-3			
24					
23					
22					
21					
20	-2				
19					
18					
17					
16	0				

Raw	2010
20	-3
19	
18	0
17	
16	
15	+1

2010	
Avg Abs Mod	1.7
Students	6

In the example above, raw scores of 20 awarded by the school were moderated down 3 points by the IB, raw scores of 18 did not change after moderation, and scores of 15 were moderated up 1 point. No student had a raw score of 19, 17, or 16. The table begins with the highest raw score awarded and ends with the lowest.

"Avg Abs Mod" stands for Average Absolute Moderation, the average number of points all students' scores were moderated, regardless of direction.

	2010	2011	2012	2013	2014
Avg Abs Mod	6.0	4.3	8.8	12.1	9.1
Students	13	8	6	15	11

# IB SCORE REPORTS

Remember, this case study is based on real data. This is thus an incredible, fascinating trend. Year after year, scores were substantially moderated down.

While we would hope that the Chemistry HL teachers and administrators at Acadamigo International were already aware of this trend, our experience tells us that they may not have been.

Unless someone is helping your teachers organize and display their data clearly, it's actually quite easy to miss important trends like this. ***Is something similar occurring at your school?***

## 2. A Remarkable Improvement in Scoring Accuracy... and Scores!

After the school received our reports and were alerted to the trend, they were able to correct it. Here is their moderation summary the following year (next page).

Looking at 2015, one can see that the highest score awarded by the school was 41, and that the IBO actually moderated that up to 42. The next highest raw score was 39, and it too was moderated up a point to 40.

Viewing the full display, we see that all of the remaining raw scores for 2015 – from 32 to 25 – were left unmoderated by the IBO.

Overall, this represents a remarkable improvement, not only in accuracy of scoring by the school, but in student performance.

Here are the highest scores after moderation for each of the six years:

<u>Year</u>	<u>Highest Score After Moderation</u>
2015	42 (41 plus 1)
2014	31 (46 minus 15)
2013	30 (48 minus 18)
2012	36 (47 minus 11)
2011	36 (42 minus 6)
2010	36 (48 minus 12)

**In summary, after receiving our reports and becoming aware of the consistent pattern of downward moderation, the very next year...**

- the school had no scores moderated down,
- their students earned their highest post-moderation score in six years, and
- for the first time, the school had several students earn a 7 on the internal assessment.



# IB SCORE REPORTS

## 3. The Connection Between Scoring Accuracy and Scores

Why would an increase in scoring accuracy correspond with an increase in student performance? There are several plausible explanations, each of which is linked with an explanation for why the scores were inaccurate in the first place.

### Possibility 1

The Internal Assessment used by the teachers did not match the criteria outlined by the IBO. The students might have been doing very good work with the assessment they were provided, but the assessment itself was lacking in some respect according to the IBO criteria, and thus the students were precluded from earning maximum points. Once the students were given a better task, scores and scoring accuracy improved. (See endnote 1.)

### Possibility 2

It might be that the assessment was fine, but that the teachers did not truly understand the scoring rubric. Thus, the teachers might have been encouraging the students in the wrong direction, or conveying to them that they had met standards when in fact they had not. This is pretty clearly a suboptimal situation for student learning, and it is easy to imagine that student performance improved once the teachers better understood the IBO's expectations and scoring rubric.

### Possibility 3

It might be that the teachers understood the rubric and the IBO's standards, but were being overly generous when scoring student work. In that case, the students were not receiving accurate feedback. As with Possibility 2, they were being led to think that they were meeting standards when in fact they were not. Once the standards were properly conveyed to students through accurate scoring, they rose to the challenge and met the higher expectations.

### Possibility 4

Another possibility is perhaps the most alarming. As many readers know, the IBO does not actually review and moderate each student's internal assessment. Instead, they select a sample of internal assessments for each subject at a school, based on a range of scores. The samples are not necessarily drawn from each class or section of each subject, nor from each teacher when the subject is being taught by multiple teachers.

The IBO expects schools to standardize their internal assessment scoring between teachers and sections. (See endnote 2.) This is important and the consequences of not standardizing are worth considering.

# IB SCORE REPORTS

Suppose two teachers are teaching Mathematics SL at a school, and unfortunately, those teachers are not in alignment with respect to their application of the IB scoring rubric, and in fact have quite different grading standards.

Let's further suppose that Teacher A – the lenient grader – is teaching two sections of the subject, and Teacher B – who grades exactly to the IBO standard – is teaching just one section of the subject.

Since Teacher A is teaching two sections and is also a lenient grader, his students' internal assessments are perhaps more likely to be selected for the top-scoring samples. Those internal assessments are then moderated down, and that moderation is applied to all students in that subject at that school with the same raw score, regardless of which teacher scored their IA.

Thus, it might be that a student whose work was properly scored by Teacher B would see her score reduced, because samples of Teacher A's scoring led to a blanket downward moderation. In this hypothetical case, the lack of standardization between teachers at a school results in a student being unfairly penalized.

## 4. Conclusions

Our goal is to help schools gain insights from their IB results to improve teaching and learning.

This case study showed that the discovery of consistent downward moderation on a particular internal assessment led to an increase in both scoring accuracy and scores the following year.

While our reports cannot identify the cause of things like downward moderation, they do uncover and illuminate such trends, and thereby serve as catalysts for investigation, reflection, and productive discussions among the teachers and administrators.

---

IB Score Reports is a service by Acadamigo. We produce beautiful, clear, informative displays of IB results that help teachers, coordinators, principals, and directors identify program strengths and weaknesses.

Visit us on the web or drop us an email. We'd love to help your school.

[support@acadamigo.com](mailto:support@acadamigo.com)

# IB SCORE REPORTS

## Endnotes

1. "...[I]nternal assessment is conducted by applying a fixed set of assessment criteria for each course.... These criteria describe the kinds and levels of skills that must be addressed in the internal assessment. Teachers should ensure that students are familiar with the internal assessment criteria and that the pieces of work chosen for use in internal assessment address these criteria effectively. This is especially significant in the group 4 experimental sciences, where the portfolio of practical work might quite legitimately contain a number of pieces of work that are not suitable for use against the assessment criteria. The group 4 internal assessment criteria are intended to address a particular set of skills that may not be evident in some standard science laboratory work." (*Diploma Programme Assessment Principles and Practice*, IBO, 2004, p. 31.)

2. "When the school entry for a given course is large enough to split into different classes and more than one teacher is involved in carrying out the internal assessment, the IBO expects these teachers to share the internal assessment and work together to ensure they have standardized between them the way in which they apply the criteria. A single moderation sample is requested from the school, which in all probability will contain candidate work marked by the different teachers involved." (*Diploma Programme Assessment Principles and Practice*, IBO, 2004, p. 41.)