İSTANBUL UNIVERSITY CERRAHPAŞA

HAYEF: Journal of Education, 19(2); 155-163

DOI: 10.5152/hayef.2022.22011

Received: February 28, 2022
Accepted: May 25, 2022

HAYEF: Journal of Education

RESEARCH ARTICLE

The Relationship Betweens Predicted Grades and Final Assessment Grades in the International Baccalaureate Diploma Program

Oğuzhan ÖZCAN[1], Akın METLİ[2], Dinçer AKIŞ[3]

[1]Department of Science, Bilkent Erzurum Laboratory School (BELS), Erzurum, Turkey
[2]SEV American College (SAC), High School Administration, İstanbul, Turkey
[3]Bilkent Erzurum Laboratory School (BELS), Erzurum, Turkey

Abstract

This study aimed to explore the relationship between teachers' predicted grades and final assessment grades of the high school students in the International Baccalaureate Diploma Program. Simple linear regression analysis was conducted to explore to what extent the teachers' estimated grades predict the final assessment grades in Turkish, English, Physics, Chemistry, Biology, and Mathematics courses in the International Baccalaureate Diploma Program. The quantitative analysis revealed that predictions were varied (from .18 to .53 for standard level and from .3 to .54 in high-level courses). Furthermore, the correlation coefficient was utilized to explain the degree to which the teachers' predicted grades are associated with the final assessment grades awarded by the International Baccalaureate. This quantitative analysis similarly revealed that the results varied from low to moderate relationships (ranging from .42 to .73 in standard level and from .17 to .73 in high-level courses). Besides, contrary to the earlier findings in the literature, the study also revealed that students were under-predicted (pessimistic grades rather than optimistic grades). These findings highlight some important implications for educators regarding the reliability of teachers' predicted grades, which may be used in the university admissions or used by the international organizations for awarding the final marks.

Keywords: Assessment, external examination, International Baccalaureate Diploma Program, predicted grade, reliability

Introduction

Teachers' judgments of students' achievement performance play a major role in shaping students' educational careers, especially during university admissions. Teacher judgments also influence many important classroom and school decisions, including instructional planning, screening, placement, referrals, and communication with parents (Gittman & Koster, 1999; Hoge, 1984; Sharpley & Edgar, 1986; Stiggins & Conklin, 1992, as cited by Martínez et al., 2009) or may have an impact on the study patterns, self-perceptions, attitudes, effort, and motivation of students (Black & Wiliam, 1998; Brookhart, 1997; Rodriguez, 2004, as cited by Martínez et al., 2009).

More specifically, teachers' predicted grades (namely, forecast, expected, or estimated grades), which are based on teacher judgments of student achievement, also influence awarding students' tuition scholarships or course exemption at the university placement decisions in programs or ability groups, grade retention, and ultimately for students' future academic pathways (Meissel et al., 2017). Moreover, predicted grades may be used to monitor student progress, serve as a motivational tool (Martinez, 2001), or used within a school for teacher accountability purposes (Gill, 2019). Therefore, given its significance, predicted grades should reflect students' learning accurately as much as possible, especially because accurately predicted grades need to be submitted to the admission offices at the universities during the application process.

Due to the fact that international university admissions start at an earlier date before students' graduation from high school, students' achievement levels in the international program examinations such as International Baccalaureate Diploma Program (IBDP) may not be taken into consideration during university admissions. Since external international examinations are not held until the end of the academic year before graduation, teachers may judge students' forecast performance in the external international examinations based on students' current performance in their mock examinations, class participation, homework, quizzes, or internal assessment assignments, and so forth. However, some critical issues arise regarding the reliability of teachers' judgments, subjectivity of teachers, or validity of teachers' summative assessments, which raise further questions about to what extent teachers' estimated grades reflect students' academic achievement in the external assessments and how this may have an influence on students' educational careers in terms of admissions, course exemptions, scholarships, or even the awarding of their international diplomas.

Since prior research studies in the literature only focused on the analysis of predicted grades in the Advanced Placement (AP) or General Certificate of Secondary Education (GCSE), the current study intends to fill in the literature gap by focusing on the reliability of teachers' predicted grades in the context of IBDP offered by the IB. Founded in 1968, the IB offers a continuum of education comprising four different programs: Primary Years Program for students aged 3–12, Middle Years

Corresponding Author: Akın METLİ **E-mail:** metli@sevkoleji.k12.tr

Cite this article as: Özcan, O., Metli, A., & Akış, D. (2022). The relationship betweens predicted grades and final assessment grades in the international baccalaureate diploma program. *HAYEF: Journal of Education.*, 19(2), 155-163.

Program for students aged 11–16, Diploma Program for students aged 16–19, and Career-related Program for students aged 16–19. The IB is missioned to create a better world through education. Specifically, the IB continuum programs aim at raising internationally minded individuals who academically excel in their studies. The IBDP, which is the main educational context for this research, is made up of six subject groups (sStudies in Language and Literature, Language Acquisition, Individuals and Societies, Sciences, Mathematics, and the Arts), offered at the standard level (SL) and high level (HL). The IBDP also includes the core, comprising theory of knowledge, creativity, activity, service, and the extended essay (IB, n.d.). Students are expected to complete some coursework (internal assessment) and sit external written examinations prepared and marked by the IB. The IB has two external examination sessions, one of which is in November and the other one in May (IB, 2018). Due to the coronavirus disease-2019 (COVID-19) pandemic, there have been school closures, restrictions on school openings, postponements, or cancellations of international examinations in the 2019–2020 academic year. The IB also decided not to hold the IBDP examinations in May 2020 due to the COVID-19 pandemic. Owing to the fact that the recent COVID-19 pandemic has brought challenges to external assessments such as the cancellation of exams in International General Certificate of Secondary Education (IGCSE) or IBDP, students would instead receive calculated grades based on their teachers' predicted grades and/or coursework/internal assessment marks, which reinforces the importance of teachers' accurate judgments of students' achievement performance.

Therefore, the aim of this study is to investigate the relationship between teachers' predicted grades and high school students' final external assessment grades in the IBDP context. Earlier research on the relationship between internal and external assessment in IBDP, undertaken by Author 1, Author 2, and Author 3 (2020), specified the need for this particular research in the recommended future studies. Hence, the current study will specifically explore to what extent the teachers' predicted grades predict the final external assessment grades in Turkish, English, Physics, Chemistry, Biology, and Mathematics courses in the IBDP at an IBDP school in eastern Turkey. Furthermore, the study will explain the degree to which the teachers' predicted grades are associated with the final external assessment grades awarded by the IB. This study will highlight some important implications for educators regarding the reliability of teachers' judgments of student performances in external international assessment practices, the impact of teachers' judgments of student performances, as well as the development of alternative assessment methods.

Literature Review

With regard to the analysis of predicted grades versus actual grades in the literature, previous research has shown that teachers' predictions of student performance on external assessments are highly inaccurate (Hayward et al., 2005; Murstein 1965). For example, Murphy and Wyness (2020) studied the UK's university application system, in which students apply based on predicted examination grades, rather than actual results. Using 3 years of UK university applications data, they found that only 16% of applicants' predicted grades are accurate, with 75% of applicants having over-predicted grades. The research by Hayward et al. (2005) also pointed out that grade predictions are notoriously inaccurate as teachers are correct in only 45% of cases. Similarly, Gill and Chang (2013) investigated the accuracy of forecast grades in Oxford, Cambridge, Royal Society of Arts (OCR) A levels in relation to the final grade. Overall, the forecast grade in 2012 was correct 48.29% of the time. Furthermore, Dhillon's (2005) analyses of required estimated grades to a British awarding body revealed that teachers' estimates of AS level grades also showed relative imprecision. In addition to being inaccurate, these estimates were more commonly optimistic than pessimistic, which was consistent with the earlier findings of the studies by Child and Wilson (2015) and Gill and Chang (2013).

In addition to these reported inaccuracies in predicted grades, the relationship between teacher judgments and measured student performance has also been inconsistent with a wide range of correlations reported: .03–.92 (Hoge & Coladarci, 1989; Südkamp et al., 2012, as cited by Meissel et al., 2017). Gill (2019) discusses that grade predictions are unreliable because of the multitude of factors which are beyond the teacher's control. He exemplifies that there may be many different reasons why students do not perform as expected. For example, there may be certain factors that influence students' performance after predictions were made in a way that students can show better than expected performance by working very hard or show a worse than expected performance by doing little work before the external examinations.

On the other hand, previous research has also found a few positive correlations between teacher judgments of student achievement and the scores the students obtain on standardized tests. For example, Murphy (1979) investigated the level of agreement between teachers' grade predictions and students' actual General Certificate in Education (GCE) grades (O and A level) for the whole sample of 1892 candidates and found that there was a reasonably high level of agreement. Machts et al. (2016) also found an average correlation of .61 between teachers' judgments of their students' cognitive abilities and their actual academic achievement. Lastly, Delap (1994) investigated the relationship between teachers' predictions of A-level grades and the actual grades and found a significant correlation between the predicted and actual grades awarded.

Yet, the strength of this association can vary considerably across subjects, grades, teachers, and classroom practices (Hoge & Coladarci, 1989; Martínez et al., 2009; Perry & Meisels, 1996). This variation may also be associated with differences in teaching experience (Leinhardt, 1983) and with differences in content area or pedagogical professional development (Rodriguez, 2004); due to teachers' emotional state or mood (Brackett et al., 2013) or their unpredictable standards (Cizek et al., 1995), so there is considerable heterogeneity in how teachers weigh the importance of different grading principles (McMillan et al., 2002).

Alternatively, teachers' assessments might be subject to bias, as well (Hoge & Butcher, 1984; Martínez et al., 2009; Meissel et al., 2017), owing to factors such as students' ethnicity (Darling-Hammond, 1995), socioeconomic status (Darley & Gross, 1983), gender (Marcenaro-Gutierrez & Vignoles, 2015; Tiedemann, 2002), and special needs (Gear, 1976; Martínez & Mastergeorge, 2002; Murphy and Wyness, 2020), school type (Marcenaro-Gutierrez & Vignoles, 2015), classroom conduct, (Bennett et al., 1993), and English for speakers of other languages status (McKown & Weinstein, 2008; Rubie-Davies et al., 2012; Südkamp et al., 2012)). That said, Snell et al. (2008) found that prediction bias is not particularly related to the gender, class, or schooling of the student but is closely linked to the predicted grades themselves—students predicted low grades perform above expectations, and vice-versa.

Considering all these challenges regarding teachers' judgments of student achievement, Anders et al. (2020) inquired whether it is possible to improve teachers' predictions using detailed measures of students' past performance and non-linear and machine learning approaches. Their study revealed that even with advanced statistical techniques, and rich achievement data, models still generate low rates of prediction accuracy, with varying rates of accuracy across student achievement, school type, and subjects. This raises the question as to why predicted grades continue to form such a crucial part of international education

systems. Since the estimated grades are not frequently accurate, Higgins and Scott (1992) (as cited by Delap, 1995) propose that higher education institutions should not place too much reliance on the estimated grades or universities placements based on estimated grades should be delayed until after the external examination results are known.

Methods

Research Design

This research utilized a quantitative method with purposive sampling to explore the relationship between the IBDP teachers' predicted grades and the final grades awarded by the IB for the IBDP candidates at a private IB authorized school. Simple linear regression analysis was conducted to explore to what extent the teachers' predicted grades predict the final assessment scores on the IB diploma. Furthermore, the correlation coefficient was used in order to explain the degree to which the teachers' predicted scores are associated with the final assessment scores of the IBDP students.

Sampling

The entire sample was drawn from high school students (n = 349, 165 males and 184 females) who attended a private national school in eastern Turkey between the years 2012 and 2019 and took the IBDP May examination session. This case school is affiliated with the Ministry of National Education and also implements international programs such as IGCSE in grades 9 and 10 and IBDP in grades 11 and 12, as mandated by the school's bylaws. The sample was a mono-cultural group as all students were Turkish and come mainly from eastern Turkey. The students' parents are on a diverse socio-economic level. All students were on either partial or full academic merit scholarships. The entire sample group took the IGCSE exams at the end of grade 10. At the end of the 2-year IGCSE program, students were admitted to the IBDP with the same selection criteria. The sample group who succeeded in the IGCSE examinations took IBDP examinations at the end of grade 12.

Data Collection

The data consisted of teachers' predicted grades and the final assessment marks of the IBDP candidates (the sample for each SL and HL course is presented in Tables 1–4) who were enrolled at the case school between the years 2012 and 2019. The year 2020 was not included because the IB had announced that the May 2020 final examinations would not take place owing to the COVID-19 pandemic. The data obtained from the years 2012–2019 consisted of the sample group's predicted grades given by the teachers in Turkish, English, mathematics, physics, chemistry, and biology courses, as well as the final assessment scores of Turkish, English, mathematics, physics, chemistry, and biology courses. The predicted grades were internally given by the case school's teachers and shared with the IB. The final assessment scores on the IB diploma were based on the marks given by the IB examiners for students' examination papers of the IBDP subjects set by the IB at the end of their study in grade 12 in the IBDP. All the data regarding teachers' predicted grades and candidates' final assessment scores were retrieved electronically from the International Baccalaureate Information Systems (IBIS) account and collected with permission from the case school. The school administration gave ethical approval to undertake and publish this current research.

Data Analysis

This study utilized quantitative methods to investigate the relationship between the teachers' predicted grades and the final assessment grades in the IBDP context. In this study, the independent variables (predictors) are teachers' predicted grades of IBDP Turkish, English, mathematics, physics, chemistry, and biology courses. The dependent variables are the final assessment scores of Turkish, English, mathematics, physics, chemistry, and biology courses. Descriptive statistics were obtained first through Statistical Package for the Social Sciences for all independent and dependent variables. Then, Pearson correlation coefficient values were examined to see the relationship between the internal assessment scores and external assessment scores in the pertinent aforementioned IBDP courses. Finally, the data were analyzed with simple linear regression to see to what extent the teachers' predicted grades scores explain the variation in the final assessment scores in the abovementioned IBDP courses.

Statistical assumptions were checked before conducting a linear regression for the prediction of dependent variables, higher level, and standard level-predicted grade scores in biology, chemistry, English, mathematics, physics, and Turkish, along with the final grades of higher-level and standard-level subjects in biology, chemistry, English, mathematics, physics, and Turkish. In order to check for normality as the first assumption, a visual inspection of histogram and the examination of skewness and kurtosis values have been conducted. Skewness values ranged from −.62 to .58, while kurtosis values ranged from −.73 to 1.205. The analysis of Kolmogorov–Smirnov test and skewness/kurtosis values provided us with information that the data show a normal distribution (Tabachnick & Fidell, 2013). The second assumption is homoscedasticity where variances should be equally distributed on the scatterplot. The results show that points are equally distributed above and below zero on the x axis and left and right sides of the zero on the y axis (Tabachnick & Fidell, 2013). Thus, we can assume that homoscedasticity was also satisfied for the data.

Table 1.
Descriptive Statistics of HL Subjects of PG Grades

	Biology	Chemistry	English	Mathematics	Physics	Turkish
n	176	206	111	132	108	194
Mean	5.187	4.791	4.936	5.083	4.787	5.762
Median	5.000	5.000	5.000	5.000	5.000	6.000
Mode	6.00	4.00	5.00	5.00	4.00	6.00
Standard deviation	.970	1.012	1.138	1.041	.947	.902
Skewness	−.271	.315	.088	.243	.575	−.283
Standard error of skewness	.183	.169	.229	.211	.233	.175
Kurtosis	−.571	−.388	−.688	−.262	−.164	−.298
Standard error of kurtosis	.364	.337	.455	.419	.461	.347
Range	4.00	4.00	4.00	4.00	4.00	4.00
Minimum	3.00	3.00	3.00	3.00	3.00	3.00
Maximum	7.00	7.00	7.00	7.00	7.00	7.00

Note: HL = high level; PG = post-graduation.

Table 2.
Descriptive Statistics of HL Subjects of Final Grades

	Biology	Chemistry	English	Mathematics	Physics	Turkish
n	176	206	111	132	108	194
Mean	5.022	4.684	5.180	5.166	4.675	6.041
Median	5.000	5.000	5.000	5.000	5.000	6.000
Mode	5.00	5.00	5.00	5.00	5.00	6.00
Standard deviation	.868	1.041	.799	.900	1.166	.673
Skewness	−.203	.114	−.013	.299	.050	−.357
Standard error of skewness	.183	.169	.229	.211	.233	.175
Kurtosis	.213	−.243	−.255	−.434	−.703	.231
Standard error of kurtosis	.364	.337	.455	.419	.461	.347
Range	5.00	5.00	4.00	4.00	5.00	3.00
Minimum	2.00	2.00	3.00	3.00	2.00	4.00
Maximum	7.00	7.00	7.00	7.00	7.00	7.00

Note: HL = high level.

Results

Descriptive statistics were used to describe basic features of the data used in the study. Results of the descriptive statistics provide basic summaries of the sample and measures that have been used to explore the relationship between teachers' predicted grades and final assessment grades of the high school students in the IBDP. The descriptive statistics results of all variables in terms of standard deviations, means, skewness, and kurtosis are presented in the descriptive statistics tables. They are presented in two different tables to indicate statistics of IBDP students who chose SL subjects and HL subjects separately in biology, chemistry, English, math, physics, and Turkish courses. Table 1 shows the descriptive statistics for the higher-level predicted grade scores in biology, chemistry, English, math, physics, and Turkish.

Table 2 shows the descriptive statistics for the higher-level final grades in biology, chemistry, English, math, physics, and Turkish.

Table 3 shows the descriptive statistics for the standard-level predicted grade scores in biology, chemistry, English, math, physics, and Turkish.

Table 4 shows the descriptive statistics for the standard-level final grades in biology, chemistry, English, math, physics, and Turkish.

Regression Analysis

The beta-coefficient values provide us the degree of change in the outcome variable for every one unit change in the predictor variable. The "β_1" value against the final scores of the IBDP standard-level courses shows slope coefficients. The gradient values are above zero for all courses, which can be tested by significance values against the IBDP final scores. Significance values presented in Tables 5 and 6 show significant evidence that the gradient values are not zero ($p < .05$). This shows that a change in IBDP predicted grades has an effect on the variations in the IBDP final grades.

The values in Table 5 show that the slope coefficient for the IBDP SL biology final grades is .832, so the IBDP SL biology predicted grades increase by .832 for each 1 point increase in IBDP SL biology final grades ($\beta = .832$; $p < .05$). The IBDP SL chemistry final grades increase by .556 for each 1 point increase in the IBDP HL chemistry predicted grades ($\beta = .556$; $p < .05$). The IBDP SL English final grades increase by .674 for each 1 point increase in the IBDP HL English predicted grades ($\beta = .674$; $p < .05$). The IBDP SL math final grades increase by .738 for each 1 point increase in the IBDP HL math predicted grades ($\beta = .738$; $p < .05$). The IBDP SL physics final grades increase by .695 for each 1 point increase in the IBDP HL physics predicted grades ($\beta = .695$; $p < .05$). The IBDP SL Turkish final grades increase by .454 for each 1 point increase in the IBDP HL Turkish predicted grades ($\beta = .454$; $p < .05$).

As for the "β_1" values in the higher-level IBDP courses, they show slope coefficients. The gradient values of all courses are above zero, which can be tested by significance values against IBDP final grades. The values in Table 6 show that a change in the IBDP higher-level courses predicted grades have an effect on the IBDP higher-level final grades. The slope coefficient for the IBDP HL biology final grades is .699, so the IBDP SL biology predicted grades increase by .699 for each 1 point increase in the IBDP SL biology final grades ($\beta = .699$; $p < .05$). The IBDP HL chemistry final grades increase by .658 for each

Table 3.
Descriptive Statistics of SL Subjects of PG Grades

	Biology	Chemistry	English	Mathematics	Physics	Turkish
n	74	40	235	218	71	155
Mean	4.864	5.325	5.246	5.445	4.943	5.625
Median	5.000	5.000	5.000	5.500	5.000	6.000
Mode	5.00	6.00	5.00	5.00	5.00	6.00
Standard deviation	.997	1.047	.826	1.163	1.054	.822
Skewness	.364	−.143	.155	−.344	−.186	−.198
Standard error of skewness	.279	.374	.159	.165	.285	.195
Kurtosis	−.306	−.775	.113	−.668	−.697	−.436
Standard error of kurtosis	.552	.733	.316	.328	.563	.387
Range	4.00	4.00	4.00	4.00	4.00	3.00
Minimum	3.00	3.00	3.00	3.00	3.00	4.00
Maximum	7.00	7.00	7.00	7.00	7.00	7.00

Note: SL = standard level; PG = post-graduation.

Table 4.
Descriptive Statistics of SL Subjects of Final Grades

	Biology	Chemistry	English	Mathematics	Physics	Turkish
n	74	40	235	218	71	155
Mean	4.918	5.000	5.140	5.605	5.253	6.103
Median	5.000	5.000	5.000	6.000	5.000	6.000
Mode	5.00	5.00	5.00	6.00	5.00	6.00
Standard deviation	.872	.960	.729	1.115	1.024	.765
Skewness	−.349	.000	.244	−.620	−.287	−.441
Standard error of skewness	.279	.374	.159	.165	.285	.195
Kurtosis	1.205	−.297	.219	−.342	−.267	−.418
Standard error of kurtosis	.552	.733	.316	.328	.563	.387
Range	5.00	4.00	4.00	4.00	4.00	3.00
Minimum	2.00	3.00	3.00	3.00	3.00	4.00
Maximum	7.00	7.00	7.00	7.00	7.00	7.00

Note: SL=standard level.

1 point increase in the IBDP HL chemistry predicted grades (β=.658; $p < .05$). The IBDP HL English final grades increase by 1.026 for each 1-point increase in the IBDP HL English predicted grades (β=1.026; p=.444). So, the IBDP higher-level English course predicted grades do not statistically predict students' performance in the IBDP higher-level English course final grades. The IBDP HL math final grades increase by .810 for each 1 point increase in the IBDP HL math predicted grades (β=.810; $p < .05$). The IBDP HL physics final grades increase by .594 for each 1 point increase in the IBDP HL physics predicted grades (β=.594; $p < .05$). The IBDP HL Turkish final grades increase by .227 for each 1 point increase in the IBDP HL Turkish predicted grades (β=.227; $p < .05$).

Furthermore, a simple linear regression model was used to analyze to what extent the IBDP predicted grades predicted the IBDP final grades in the standard-level biology, chemistry, English, mathematics, physics, and Turkish courses. As seen in Table 7, 53% (R^2= *.53;* $R^2_{adjusted}$ =*.53*) of the variation in IBDP standard-level biology final grades can be explained by the model including only IBDP SL biology predicted grades. A total of 26% (R^2= *.26;* $R^2_{adjusted}$ = *.26*) for the IBDP SL chemistry predicted grades; 35% (R^2= *.35;* $R^2_{adjusted}$ =*.35*) for the IBDP SL English predicted grades; 50% (R^2= *.50;* $R^2_{adjusted}$ = *.50*) for the IBDP SL mathematics predicted grades; 46% (R^2= *.46;* $R^2_{adjusted}$ = *.46*) for IBDP SL physics predicted grades; 18% (R^2= *.18;* $R^2_{adjusted}$ = *.18*) for IBDP SL Turkish predicted grades can be explained by this model.

Another simple linear regression model was used to analyze to what extent the IBDP predicted grades predict the IBDP final grades in higher-level biology, chemistry, English, mathematics, physics, and Turkish courses. As seen in Table 8, 39% (R^2= *.39;* $R^2_{adjusted}$ = *.39*) of the variation in the IBDP higher-level biology final grades can be explained by the model including only the IBDP higher-level biology predicted grades. A total of 46% (R^2= *.46;* $R^2_{adjusted}$ = *.46*) for the IBDP higher-level chemistry predicted grades; 52% (R^2= *.52;* $R^2_{adjusted}$ = *.52*) for the IBDP higher-level English predicted grades; 49% (R^2= *.49;* $R^2_{adjusted}$ = *.49*) for the IBDP higher-level mathematics predicted grades; 54% (R^2= *.54;* $R^2_{adjusted}$ =*.54*) for the IBDP higher-level physics predicted grades; 3% (R^2= *.03;* $R^2_{adjusted}$ = *.03*) for the IBDP higher-level Turkish predicted grades can be explained by this model.

Table 5.
Unstandardized and Standardized Regression Coefficients for IBDP SL Courses

	Unstandardized Coefficients		Standardized Coefficients		
Model	*B*	Standard Error	β	*t*	Significance
Constant	.772	.462		1.671	.000
Biology SL final	.832	.217	.727	8.993	.000
Constant	2.547	.774		3.290	.002
Chemistry SL final	.556	.152	.510	3.652	.001
Constant	1.781	.310		5.749	.000
English SL final	.674	.060	.595	11.301	.000
Constant	1.309	.287		4.566	.000
Math SL final	.738	.050	.707	14.707	.000
Constant	1.294	.489		2.646	.010
Physics SL final	.695	.091	.675	7.600	.000
Constant	2.857	.484		5.899	.000
Turkish SL final	.454	.079	.422	5.761	.000

Note: IBDL=International Baccalaureate Diploma Program; SL=standard level.

Table 6.
Unstandardized and Standardized Regression Coefficients for IBDP HL Courses

	Unstandardized Coefficients		Standardized Coefficients		
Model	*B*	Standard Error	β	*t*	Significance
Constant	1.675	.337		4.971	.000
Biology HL final	.699	.066	.626	10.579	.000
Constant	1.708	.240		7.114	.000
Chemistry HL final	.658	.050	.677	13.153	.000
Constant	−.380	.495		−.769	.444
English HL final	1.026	.094	.721	10.874	.000
Constant	.897	.379		2.365	.020
Math HL final	.810	.072	.701	11.208	.000
Constant	2.009	.259		7.750	.000
Physics HL final	.594	.054	.732	11.046	.000
Constant	4.292	.579		7.589	.000
Turkish HL final	.227	.095	.170	2.384	.018

Note: IBDL=International Baccalaureate Diploma Program; HL=high level.

Table 7.
Regression-Adjusted R^2 *Results for IBDP SL Courses*

	Biology Final	Chemistry Final	English Final	Mathematics Final	Physics Final	Turkish Final
Biology PG	.53					
Chemistry PG		.26				
English PG			.35			
Math PG				.50		
Physics PG					.46	
Turkish PG						.18

Note: PG=post-graduation; IBDL=International Baccalaureate Diploma Program; HL=high level.

Table 8.
Regression-Adjusted R^2 *Results for IBDP HL Courses*

	Biology Final	Chemistry Final	English Final	Mathematics Final	Physics Final	Turkish Final
Biology PG	.39					
Chemistry PG		.46				
English PG			.52			
Math PG				.49		
Physics PG					.54	
Turkish PG						.03

Note: PG=post-graduation; IBDL=International Baccalaureate Diploma Program; HL=high level.

Correlation Analysis

The study found that there is a statistically moderate relationship between IBDP higher-level biology, chemistry, English, math, physics predicted grades, and final grades. The Pearson correlation analysis showed a statistically weak relationship between the IBDP higher-level Turkish predicted grades and final grades, as shown in Table 9.

The analysis also showed that there is a statistically moderate relationship between the IBDP standard-level biology, chemistry, English, math, physics-, Turkish predicted grades, and final grades, as shown in Table 10.

Discussion, Conclusion and Recommendations

The data obtained from the case study school indicate that the strength of the prediction of the IBDP predicted grades for final grades determined after external exams is lower than expected. It is interesting to note that earlier studies from the literature showed that the vast majority of students are over-predicted, and they are given optimistic predicted grades (Everett and Papageorgiou, 2011; Hayward et al., 2005). On the other hand, the results of the current study indicated that in the case study school, the majority of students are under-predicted. There are different factors stated in the literature explaining the discrepancy between predicted grades and achieved grades, such as socio-economic status, ethnicity, gender, school type, teacher turnover rate, teacher experience, and accountability of teachers to the institution and family (Baird, 1997; Delap, 1994; Dhillon, 2005; Murphy and Wyness, 2020).

The case study school accepts students via a two-stage entrance exam. At the end of grade 10, they take the IGCSE exams, and they must get at least a "C" or above in physics, chemistry, biology, math, Turkish, and English in order to be qualified to continue for the IBDP. So, it can be said that the student group is almost homogeneous in terms of their academic abilities. As explained in the sampling section, the group is mono-cultural (Turkish) and mostly comes from eastern

Table 9.
Correlations for IBDP HL Courses Predicted Grades and Final Grades

	Biology Final	Chemistry Final	English Final	Mathematics Final	Physics Final	Turkish Final
Biology PG	.63					
Chemistry PG		.68				
English PG			.72			
Math PG				.70		
Physics PG					.73	
Turkish PG						.17

Note: PG=post-graduation; IBDL=International Baccalaureate Diploma Program; HL=high level.

Table 10.
Correlations for IBDP SL Courses Predicted Grades and Final Grades

	Biology Final	Chemistry Final	English Final	Mathematics Final	Physics Final	Turkish Final
Biology PG	.73					
Chemistry PG		.51				
English PG			.60			
Math PG				.71		
Physics PG					.68	
Turkish PG						.42

Note: PG=post-graduation; IBDL=International Baccalaureate Diploma Program; HL=high level.

Turkey and has a diverse socioeconomic status. Thus, it is difficult to state that ethnicity has had an impact on grade prediction accuracy for this case study school. Totally 90% of the students have full academic merit scholarships, and the remaining group has partial academic merit scholarships in the school, so it is also difficult to claim that socioeconomic status plays an important role to explain the discrepancy between predicted grades and final grades. It seems that other factors such as teacher experience, teacher turnover rate, type of subject, and so on may have a greater impact at this point.

Studies investigating this inconsistency mainly focus on GCSE and A-level qualifications, but the authors of this current study present one of the very first studies focusing on the discrepancy between the IBDP predicted and final grades regarding the higher-level and standard-level courses separately. Students sit examinations and complete coursework—with a different weight for different subject groups—to earn their IB diploma (IB, 2018). Physics, chemistry, and biology are in the same group and have the same weightages for exams and coursework. When it comes to predictions, the results of this study indicated that the correlations between predicted grades and final grades are moderate and close to each other, especially for HL students in these science courses. Additionally, when English and mathematics HL courses are considered, their correlations between predicted grades and final grades are also moderate and close to the science subjects. In the light of these results, it can be claimed that having an almost homogeneous group of students may prevent teachers from being prejudiced while giving predicted grades.

The results of the study revealed another interesting point about the predictions regarding the Turkish course. Based on the degree of prediction of the IBDP Turkish predicted grades for final grades (HL course: $R^2 = .03$; $R^2_{adjusted} = .03$ and SL course: $R^2 = .18$; $R^2_{adjusted} = .18$), students usually are under-predicted in the Turkish course. Contrary to the findings of a study which claimed that "subjects in which there is a high proportion of A-grades predicted would also have a large percentage of over-predictions" (Snell et al., 2008); in the case study school, Turkish is the course in which the largest number of students—with respect to all other courses—get a 7 as a final grade—the highest grade in IBDP—but according to the data obtained from this study, Turkish is the course in which students are under-predicted the most. An explanation for this could be teachers' judgments which are underestimating the potential of their students. Another explanation for this discrepancy could be the differences between the weightages of coursework across the IBDP subjects (Snell et al., 2008). The weightage of coursework is 40% in Turkish SL, 30% in Turkish HL, and 20% in science and math courses. Academic coursework in IBDP is internally assessed and externally moderated, so the grade given by the teacher can be changed by the moderator, which may have an impact on the inconsistency between predicted and final grades. It should also be noted that increasing weightage of academic coursework may widen the gap between predicted and final grades by increasing the impact factor of disagreement between teachers and moderators over the grading of coursework.

The IBDP is a 2-year international program. Students receive their predicted grades according to the performance they exhibit over these 2 academic years. Having different teachers in the first year and second year may limit the ability of teachers to make accurate judgments about their students' potential abilities and achievement levels because in this way teachers may get the exam or quiz results from the first year to make an accurate prediction. However, evaluating other characteristics such as perseverance, enthusiasm, creativity, and being innovative may be more difficult for the new teachers. Hence, schools that have a high teacher turnover rate should make effective and sustainable planning to help new teachers for accurate and confident predictions (Ronfeldt et al., 2013).

Teachers may think that giving a predicted grade lower than they expect can encourage students to work harder. If we consider this as one of the factors explaining the inconsistency between predicted and final grades, lower predicted grades may trigger some students to study harder with increased motivation. Besides, it should be added that most of the students work harder during the time between predicted grades and external exams, but if they are already given high predicted grades, they may think that it is meaningless to put more effort because places at universities are granted with these high predicted grades. Unfortunately, teachers cannot evaluate and reflect on the performance of their students while giving their predicted grades (Snell et al., 2008). From this point of view, the hard work of students before and during the exam period may be considered as one of the factors that has an impact on the discrepancy between predicted and final grades (Murphy and Wyness, 2020).

Challenging circumstances that are experienced such as COVID-19 make it even more difficult for teachers to make accurate predictions because time spent on face-to-face lessons is limited and it is difficult to observe the learning characteristics of students during virtual education period. Teachers should get to know students better via using formative assessment tools to give predicted grades with high accuracy and confidence, but virtual lessons decrease the possibility of implementing effective formative assessment techniques, especially for teachers who have difficulty in using technology and educational applications. At this point, educators need to be aware that instead of reflecting students' true potential abilities with high accuracy, predicted grades can be the mirror images of teachers' beliefs about their students. Along with the academic strength of students, appearance, in-class behaviors, and personal relations with teachers can also shape teachers' beliefs and cause biased predictions about students' academic achievement levels (Macionis & Plummer, 2012).

It is clear that giving 100% correct predicted grades by teachers is impossible, but if these predictions are going to keep their important role in the university admission process, then organizations, universities, policymakers, and teachers should work together in order to find ways and develop new strategies to increase the accuracy of predicted grades (Anders et al., 2020). From this point of view, institutions can make a standardization by assigning reliability coefficients according to schools' reputation about how accurately they give predicted grades, and these coefficients should be updated every year in accordance with the consistency between predicted grades and final grades. This system may encourage schools to develop better grade prediction policies. For example, schools can provide professional development training sessions for teachers about how to increase grade prediction accuracy. Alternatively, the components used by teachers to assign predicted grades can be standardized by schools so that the same criteria would be considered for a fairer evaluation of students.

Furthermore, some universities and colleges start their application process at the end of December, when students have a long time for further academic development, so teachers are pressured to make early predictions. However, early predictions may lead teachers to be optimistic by considering the time ahead of students for further development and cause over-predictions (Murphy and Wyness, 2020). Particularly, the IB Organization expects teachers to enter their predicted grades into the IBIS by the end of April. So, it becomes difficult for teachers to make a change in their predicted grades for students who received their predicted grades for early university applications because, at this point, a decrease or an increase may get strong reactions from students and their families, so in order not to be held responsible and not to be blamed, teachers may hesitate to reflect the change in students' performance—either negative or positive—into their finalized predicted grades and prefer to keep them as they were,

which can be another source of the discrepancy between predicted and final grades. The IBDP November session exams may be a good solution for this, because students can learn their achieved grades in the first week of January, and they can make their university applications by using their final grades rather than using predicted grades.

To conclude, when the earlier studies from the literature are considered, there is a remarkable difference between teachers' predictions and final grades for years. This fact is known by universities, colleges, and other institutions, but they keep using these grades as a critical part of their student admission system because predicted grades provide useful information about students' potential abilities and academic achievement. Moreover, COVID-19 as an extraordinary situation also increased the importance of predicted grades because the cancellation of exams limits the alternative ways for institutions to make fair decisions about admission. Thus, educational institutions such as universities and colleges rely more on predicted grades. From this point of view, it becomes clearer that policymakers, educators, and teachers should continue to work toward increasing the accuracy of predicted grades as this study specifically revealed that we cannot fully rely on the predicted grades of teachers to a great extent.

Finally, the current study has some limitations regarding the data which were collected from only one case study school in which the student group is almost homogeneous in terms of their academic abilities, being mono-cultural (Turkish), mostly being from eastern Turkey. Researchers who plan to conduct similar studies are therefore recommended to collect data from different IBDP schools located in different countries to observe the situation regarding the predicted grade—final grade discrepancy in different educational contexts and to provide an enlarged perspective. In addition, the IBDP offers a large set of courses, but only six of them were taken into account in this study. Hence, future studies are recommended to focus on different courses to see the effect of type of subject on predicted grade—final grade discrepancy by comparing their results with the results of this study. Moreover, it is recommended that future studies may also focus on the factors such as school type, teacher turnover rate, teacher experience, accountability of teachers to the institution and family, and types of subjects, which may have a role in explaining the variation between predicted and final grades, to have a deeper understanding on the phenomenon under investigation.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – A.M.; Design – O.Ö.; Supervision – A.M.; Data Collection and/or Processing – D.A.; Analysis and/or Interpretation - D.A., O.Ö.; Literature Review – A.M.; Writing Manuscript – A.M., O.Ö., D.A., Critical Review – A.M., O.Ö., D.A.

Declaration of Interests: The authors declare that they have no competing interest.

Funding: The authors declared that this study has received no financial support.

References

Anders, J., Dilnot, C., Macmillan, L., & Wyness, G. (2020). *Grade Expectations: How well can we predict future grades based on past performance?* Working Paper no. 20-14. Centre for Education Policy and Equalizing Opportunities (CEPEO): Oxford Brookes Business School.

Baird, J. H. (1997). *Teachers' estimates of A-level performance. Internal report RAC/763*. Guildford: Associated Publishers Examining Board, Mimeo.

Bennett, R. E., Gottesman, R. L., Rock, D. A., & Cerullo, F. (1993). Influence of behavior perceptions and gender on teachers' judgments of students' academic skill. *Journal of Educational Psychology*, *85*(2), 347–356. [CrossRef]

Brackett, M. A., Floman, J. L., Ashton-James, C., Cherkasskiy, L., & Salovey, P. (2013). The influence of teacher emotion on grading practices: A preliminary look at the evaluation of student writing. *Teachers and Teaching*, *19*(6), 634–646. [CrossRef]

Child, S., & Wilson, F. (2015). *An investigation of A level teachers' methods when estimating student grades. Cambridge assessment internal report*. Cambridge, UK: Cambridge Assessment.

Cizek, G. J., Fitzgerald, S. M., & Rachor, R. A. (1995). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment*, *3*(2), 159–179. [CrossRef]

Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, *44*(1), 20–33. [CrossRef]

Darling-Hammond, L. (1995). Equity issues in performance-based assessment. In M. T. Nettles & A. L. Nettles (Eds.), *Equity and excellence in educational testing and assessment* (pp. 89–114). Boston: Kluwer Publishers.

Delap, M. R. (1994). An investigation into the accuracy of A-level predicted grades. *Educational Research*, *36*(2), 135–148. [CrossRef]

Delap, M. R. (1995). Teachers' estimates of candidates' performances in public examinations. *Assessment in Education: Principles, Policy and Practice*, *2*(1), 75–92. [CrossRef]

Dhillon, D. (2005). Teachers' estimates of candidates' grades: Curriculum 2000 Advanced Level Qualifications. *British Educational Research Journal*, *31*(1), 69–88. [CrossRef]

Everett, N., & Papageorgiou, J. (2011). *Investigating the accuracy of predicted a level grades as part of 2009 UCAS admission process.* London: Department for Business, Innovation and Skills.

Gear, G. H. (1976). Accuracy of teacher judgment in identifying intellectually gifted children: A review of the literature. *Gifted Child Quarterly*, *20*(4), 478–490. [CrossRef]

Gill, G. (2019). Methods used by teachers to predict final A Level grades for their students. *Research Matters*, *28*, 33–42.

Gill, T., & Chang, Y. (2013). *The accuracy of forecast grades for OCR A levels in June 2012*. Statistics Report Series No.64. Cambridge Assessment, Cambridge: UK.

Hayward, G., Sturdy, S., & James, S. (2005). *Estimating the reliability of predicted grades*. UCAS. Retrieved from www.ucas.ac.uk/candq/grades.pdf.

Hoge, R. D., & Butcher, R. (1984). Analysis of teacher judgements of pupil achievement levels. *Journal of Educational Psychology*, *76*(5), 777–781. [CrossRef]

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, *59*(3), 297–313. [CrossRef]

International Baccalaureate. (2018). *Assessment principles and practices – Quality assessments in a digital age*. Bethesda, MD: International Baccalaureate Organization.

International Baccalaureate. (IB) (2021). *Diploma Programme*. Retrieved from https://www.ibo.org/programmes/diploma-programme/.

Leinhardt, G. (1983). Novice and expert knowledge of individual students' achievement. *Educational Psychologist*, *18*(3), 165–179. [CrossRef]

Machts, N., Kaiser, J., Schmidt, F. T. C., & Möller, J. (2016). Accuracy of teachers' judgments of students' cognitive abilities: A meta-analysis. *Educational Research Review*, *19*, 85–103. [CrossRef]

Macionis, J. J., & Plummer, K. (2012). *Sociology: A global introduction* (5th ed). Harlow: Pearson Education Limited.

Marcenaro-Gutierrez, O., & Vignoles, A. (2015). A comparison of teacher and test-based assessment for Spanish primary and secondary students. *Educational Research*, *57*(1), 1–21. [CrossRef]

Martínez, J. F., & Mastergeorge, A. (2002). Rating performance assessments of students with disabilities: A generalizability study of teacher bias. Paper presented at the Annual meeting of the American Educational Research Association. New Orleans, LA.

Martínez, J. F., Stecher, B., & Borko, H. (2009). Classroom assessment practices, teacher judgments, and student achievement in mathematics: Evidence from the ECLS. *Educational Assessment*, *14*(2), 78–102. [CrossRef]

Martinez, P. (2001). *Great expectations: Setting targets for students*. London, UK: Learning Skills Development Agency.

McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context, and the achievement gap. *Journal of School Psychology*, *46*(3), 235–261. [CrossRef]

McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *Journal of Educational Research*, *95*(4), 203–213. [CrossRef]

Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education*, *65*, 48–60. [CrossRef]

Metli, A., Özcan, O., & Akış, D. (2020). *Unveiling the relationship between internal assessment and external assessment in the International Baccalaureate Diploma Program*. Manuscript submitted for publication.

Murphy, R. J. L. (1979). Teachers' assessments and GCE results compared. *Educational Research*, *22*(1), 54–59. [CrossRef]

Murphy, R., & Wyness, G. (2020). Minority report: the impact of predicted grades on university admissions of disadvantaged groups [Minority Report]. *Education Economics*, *28*(4), 333–350. [CrossRef]

Murstein, B. I. (1965). The relationship of grade expectations and grades believed to be deserved to actual grades received. *Journal of Experimental Education*, *33*(4), 357–362. [CrossRef]

Perry, N. E., & Meisels, S. J. (1996). *How accurate are teacher judgements of students' academic performance?* Working Paper no. 96-08. Washington, DC: National Center for Educational Statistics.

Rodriguez, M. C. (2004). The role of classroom assessment in student performance on TIMSS. *Applied Measurement in Education*, *17*(1), 1–24. [CrossRef]

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement? *American Educational Research Journal*, *50*(1), 4–36. [CrossRef]

Rubie-Davies, C. M., Peterson, E. R., Flint, A., Garrett, L., Donald, L. M., Watson, P., & O'Neill, H. (2012). Ethnicity and teacher expectations in New Zealand. *Procedia – Social and Behavioral Sciences*, *69*, 256–261. [CrossRef]

Snell, M., Thorpe, A., Hoskins, S., & Chevalier, A. (2008). Teachers' perceptions and A-level performance: Is there any evidence of systematic bias? *Oxford Review of Education*, *34*(4), 403–423. [CrossRef]

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*(3), 743–762. [CrossRef]

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (7th ed). New York: Pearson.

Tiedemann, J. (2002). Teachers' gender stereotypes as determinants of teacher perceptions in elementary school mathematics. *Educational Studies in Mathematics*, *50*(1), 49–62. [CrossRef]