DOI: 10.34056/aujef.1091693

Uluslararası Bakalorya Diploma Programında İç Değerlendirme ve Dış Değerlendirme Arasındaki İlişkinin Araştırılması

Unveiling the Relationship between Internal Assessment and External Assessment in the International Baccalaureate Diploma Program

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

Article Type1: Research Article

Application Date: 23.03.2022 **Accepted Date:** 26.04.2023

To Cite This Article: Metli A., Ozcan O., & Akis D. (2023). Unveiling the relationship between internal assessment and external assessment in the International Baccalaureate Diploma Program. *Anadolu Üniversitesi Eğitim Fakültesi Dergisi (AUJEF), 7*(2), 386-403

ÖZ: Bu çalışma, Uluslararası Bakalorya Diploma Programı (IBDP) lise öğrencilerinin iç değerlendirme (IA) ve dış değerlendirme (EA) puanları arasındaki ilişkiyi araştırmayı amaçlamıştır. Yöntemler: Türkçe, İngilizce, Fizik, Kimya, Biyoloji ve Matematik derslerinde iç değerlendirme puanlarının dış değerlendirme puanlarını ne ölçüde yordadığını araştırmak için regresyon analizi yapılmıştır. Bulgular: Nicel analiz, IBDP dış değerlendirme puanlarındaki varyasyonun yalnızca küçük bir yüzdesinin, IBDP standart düzey (SL) ve yüksek düzey (HL) derslerinde yalnızca iç değerlendirme puanlarını içeren modelle açıklanabileceğini ortaya koymuştur. Ayrıca, iç değerlendirme puanlarının dış değerlendirme puanları ile ne derece ilişkili olduğunu açıklamak için korelasyon katsayısından yararlanılmıştır. Bu analiz, yüksek seviye İngilizce ve standart seviye biyoloji dersinde iç değerlendirme ve dış değerlendirme arasında orta düzeyde bir ilişki olduğunu, standart seviye ile yğksek seviye biyoloji, kimya, matematik, fizik, İngilizce ve Türkçe derslerinde iç değerlendirme ve dış değerlendirme arasında zayıf bir ilişki olduğunu ortaya koymuştur. Değer: Bu araştırma, IBDP değerlendirmesinde nihai bir not oluşturmak için bir araç olarak iç değerlendirme tutarlılığı ve güvenilirliği ve iç değerlendirmenin öğrencilerin genel akademik başarısını tahmin etmede oynadığı rol hakkında eğitimciler için bazı önemli çıkarımları vurgulamaktadır.

Anahtar sözcükler: Tutarlılık, güvenilirlik, iç değerlendirme, dış değerlendirme, Uluslararası Bakalorya Diploma Programı

1 Dr., Teacher, SEV American College (SAC), metli@sevkoleji.k12.tr, ORCID: 0000-0002-7117-9189
2 Dr., Teacher, Bilkent Erzurum Laboratory School (BELS), oguzhanozcan@bels.bilkent.edu.tr, ORCID: 0000-0002-8338-4606
3MA.,Teacher, Bilkent Erzurum Laboratory School (BELS), akis@bels.bilkent.edu.tr, ORCID: 0000-0001-6274-7992
4This research was designed as a study containing only quantitative data. The data do not include any personal information. All the necessary permissions of the school administration were taken.

ABSTRACT: This study aimed to explore the relationship between internal assessment (IA) and external assessment (EA) scores of the high school students in the International Baccalaureate Diploma Program (IBDP). Methods: Simple linear regression analysis was conducted to explore to what extent the IA scores predict the EA scores in the Turkish, English, Physics, Chemistry, Biology and Mathematics courses. Findings: The quantitative analysis revealed that only a small percentage of the variation in IBDP EA scores can be explained by the model containing only IA scores in the IBDP standard level (SL) and high level (HL) courses. Furthermore, the correlation coefficient was utilized to explain the degree to which the IA scores are associated with the EA scores. This analysis revealed that there is a moderate relationship between IA and EA in HL English and SL biology course, but a weak relationship between IA and EA in SL and HL biology, chemistry, mathematics, physics, English (only SL) and Turkish courses. Value: This research highlights some important implications for educators about the consistency and reliability of IA as a tool for forming a final mark in the IBDP assessment and the role the IA plays in predicting the students' general academic achievement.

Keywords: Consistency, coursework, reliability, external assessment, internal assessment, International Baccalaureate Diploma Program

1. INTRODUCTION

Internal assessment, or in other words coursework, has been included in various international secondary education programs such as Advanced Placement (AP), International General Certificate of Secondary Education (IGCSE) or International Baccalaureate Diploma Program (IBDP) to give students the opportunity to take responsibility in their own learning, improve their communication skills and get engaged in independent research (Bullock, et al., 2002). Furthermore, internal assessment (coursework) has also been introduced to facilitate the assessment of positive achievement (Tattersall, 1994) as well as to measure skills which could not be assessed through formal written exams (Elwood, 1999). Therefore, acting as a form of formative assessment internal assessment has been valued not only for scaffolding students for their learning processes, but also for judging their general academic performance. In that, internal assessment performance is weighted in the students' final scores of external assessments in some international education programs such as the International Baccalaureate (IB).

Founded in 1968, the IB offers a continuum of education comprising of four different programs, namely, Primary Years Program (PYP) for students aged 3-12, Middle Years Program (MYP) for students aged 11-16, Diploma Program (DP) for students aged 16-19 and Career-related Program (CP) for students aged 16-19. The mission of the IB education is to create a better world through education. The IB programs aim at developing internationally-minded individuals, encouraging personal and academic achievements and challenging students to excel in their studies and their personal development. The DP, which is the main educational context for this research, is made up of six subject groups (Studies 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 DP also includes the DP core, comprising theory of knowledge, creativity, activity, service and the extended essay (IB, n.d.). The IB has two main assessment models in the IBDP, which are namely the external assessment and internal assessment. External assessment is "a form of assessment in which question papers, assignments and tasks are specified by the awarding organisation, then taken under specified conditions … and marking or assessment judgements are made by the awarding organization" (DfE, 2015a, 2018, as cited by Vitello and Williamson, 2017). Internal assessment means that "the school (typically the candidate's teacher) that marks the candidate's work, and the IB then checks that the teacher has correctly applied the global standard through a process of moderation" (International Baccalaureate, 2018).

Internal assessment (coursework) has played a more significant role in the overall judgment of students' general academic achievement in the international programs due to the Covid-19 (coronavirus) pandemic. The Covid-19 (coronavirus) pandemic has influenced many educational systems worldwide, leading to school closures, or postponement or cancellation of international examinations taken by millions of students in more than 140 countries worldwide. Therefore, alternatively many educational systems have had to implement virtual education or online remote assessments which have been mainly based upon coursework, classroom-based assessment or internal assessment as formative assessments rather than summative assessments. In this world of rapid, and unpredictable change and challenging times due to the Covid-19 (coronavirus) pandemic, the International Baccalaureate (IB) decided to cancel its 2020 final external examinations. The IB decided to determine students' achievement based on students' internal assessment/coursework internally supervised by their teachers and externally moderated by the IB. Now, the internal assessment has been even more important than before in awarding the students' final marks on their diploma courses. Conceivably, as the current pandemic may continue for a while or possible other pandemics may be faced again in the world (WHO, 2020), educators and

researchers should be more agile to investigate the role and importance of other alternative assessments such as the internal assessments/coursework rather than external assessments for the award of final marks on students' international certificates or diplomas.

Yet, though assessment has been the most important and controversial area educators and policymakers have been investigating in educational research for a long time (Ziderman, 1984), there is still scarce research on the relationship between internal and external assessment in the international secondary education context in the current literature. While many prior studies have evaluated the reliability of written examinations, relatively little has been done to quantify the reliability of internal assessment (Benton, 2016). Some research in the literature merely focus on the Advanced Placement (AP) or General Certificate of Secondary Education (GCSE) or extensively about the coursework and end of year achievement levels at the university level, but none on the International Baccalaureate Diploma Program (IBDP), which is another widespread international education program.

Owing to the lack of research on the relationship between internal and external assessment in the IBDP context, with the intention to fill in this literature gap, the current study aimed to explore the relationship between internal and external assessment scores of the high school students in the International Baccalaureate Diploma Program (IBDP) context, specifically standard and high level Turkish, English, Physics, Chemistry, Biology and Mathematics courses in the IBDP. This research highlights some important implications for educators about the consistency of internal assessment as a tool for forming a final mark in the IBDP assessment and the role the internal assessment plays in predicting the students' general academic achievement and performance.

1.1. Literature Review

The use of internal assessment has been a matter of educational debate over the last 30 years (Torrance, 2018). There are also different views of internal assessment around the world (International Baccalaureate, 2018). On the one hand, internal assessment as a means of assessing many competencies has been an integral formative assessment component for judging students' general academic performance from elementary education to tertiary education (Badval & Sharma, 2020). In addition, a variety of assessment tools such as coursework or classroom-based assessments are considered to reduce the potential for inequity in assessment (Brown, 2002; Linn, 1992). From the perspective of learners, coursework encourages the development of important skills including creativity, communication, reflective thinking, designing, and conducting an investigation, and motivating students (Crisp & Green, 2013), independent learning, cultural diversity (International Baccalaureate, 2018) as well as a broader perspective of internationalism (Brown, 2002). With a view to analyze the value of coursework, Kandamby's study (2017) specifically showed that coursework supported promoting effective learning in a way that students obtained higher marks for their final exams with the help of coursework.

On the other hand, although there are certain learning benefits of coursework/internal assessment, there have also been some concerns regarding the consistency or reliability of coursework as a means of assessment toward the final assessment of students' academic performance. Richardson (2015) discusses that assessment by coursework alone or by a mixture of coursework and examinations tends to yield higher marks than assessment by examinations alone, which leads to debates on consistency, reliability, and validity of teacher judgments on student achievement and attainment. Some of these concerns were illustrated by Scott (1991) who focused on a research project about coursework in GCSE. In Scott's study, coursework practices were found to differ, with variation in the timing of coursework, where it

was being completed, the type of exercises that students were doing, the amount and quality of teacher input, the availability of resources, and the extent of parental and other help. All these were identified as factors acting to decrease the examination's reliability as a testing device. Furthermore, teaching to the test has been an ongoing concern not only for external assessments, but also for internal assessment (Torrance, 2018). Torrance further adds that the extent to which teachers can reliably assess their own students in the coursework or internal assessments is also uncertain. Other reported problems in the literature were also noted such as authenticity issues (academic honesty) (Barrance, 2019; Tattersall, 1994), overly formulaic tasks (Crisp & Green, 2013), the fairness in the assessment of collaborative tasks and the workload burden for students and teachers (International Baccalaureate, 2018).

The research on the relationship between coursework/internal assessment and external assessment is much more extensive in the higher education context in the literature but yielded different results or findings on the matter. For example, Bhadwal and Kishor (2012) studied the relationship between marks obtained by university students in theory papers and corresponding internal assessment scores and found out that the coefficients of correlation between internal and external assessment scores came out to be significant. Similarly, Ekolu (2016) studied the relationship between results of formative and summative assessments in the engineering education context. This study revealed that there is a strong direct relationship between semester results and final marks by students. However, these findings were not consistently similar in the literature. For instance, Murdan (2005) investigated the relationship between examination and coursework performances of university students in the Pharmacy program after observing that a significant number of pharmacy students achieved high coursework marks during a semester but performed poorly in the end-of-semester examinations. Contrary to the earlier findings, Murdan's study revealed that coursework marks were found to be higher than examination marks, in that there was only a weak correlation between coursework and examination marks of students.

Coursework assessment scores were found to be higher than external assessment scores in some other further studies in the higher education context, as well. In their study on the mark distributions and marking practices in the higher education context, both Simonite (2003) and Yorke, Bridges, and Woolf (2000) discussed that students tended to obtain better marks in coursework than in examinations. Yorke et al. also noted that the difference between coursework and examination marks tended to be greater in some disciplines (e.g., computer studies) than in others (e.g., history). Similarly, Bridges, et al., (2002) conducted a study to explore whether the standard of undergraduate student performance in coursework is higher than the performance achieved in formal exams in six subjects at four UK universities. The study showed that in English and History coursework performances are slightly higher, while, in Biology, Business Studies, Computer Studies and Law, coursework performances are higher by as much as two-thirds of one honours class.

As aforementioned, prior research on the relationship between internal and external assessment is limited in the secondary education context and only focuses on GCSE or AP contexts in the international secondary education programs, but none on the IB, which is implemented at over 5000 schools in around 160 countries worldwide since 1960s (IB, n.d.). In the GCSE context, Benton (2016) examined the overall reliability of coursework by comparing its value to that of written examinations taken at the same time in predicting future examination scores for the History and English Literature courses in the General Certificate of Education Advanced level (GCE A level) assessments in England. Analysing several years of assessment data, he found out that coursework is often just as predictive as externally marked tests in forecasting future performance. Thus, he discusses that "coursework scores must be at least fairly reliable

and strengthens the case for coursework as a valid form of assessment" (p.9). In the AP context, Wyatt, Jagesic and Godfrey (2018) examined the course performance of AP exam takers who placed out of an introductory college course as they have completed rigorous coursework in high school to students who did not take the AP exam and completed the introductory course at their respective institutions. The results indicated that AP exam takers, on average, earn the same or statistically significant higher grades in many different courses within the same subject area than do students completing the equivalent introductory coursework at their institution. Jagesic and Wyatt (2018) indicated that AP exam scores are valid indicators for course credit and placement decisions at higher education. However, in the West African Examinations context, the findings about these positive relationships were not consistent as Awomolo (1992) found that there was a low correlation between internal and external assessment in English and mathematics courses in the senior school certificate examination programs.

2. METHOD

All the necessary permissions and consents were taken both from the school administration of the case study school.

2.1. Research Design

This research utilized a quantitative method with a purposive sampling to explore the relationship between the internal assessment and external assessment scores of the students who studied the IBDP in a private IB authorized school. Simple linear regression analysis was conducted to explore to what extent the internal assessment scores as the independent variable predict the external assessment scores as the dependent variable. Furthermore, the correlation coefficient was used in order to explain the degree to which the internal assessment scores are associated with the external assessment scores of the students who studied the IBDP.

2.2. Sampling

The sample was drawn from high school students (N = 338, 164 male and 174 female) who attended a private national school between the years 2012 and 2019. This case school located in a city in eastern Turkey is affiliated with the Ministry of National Education (MoNE). The school also follows compulsory international programs such as International General Certificate of Secondary Education (IGCSE) in Grade 9 and 10 and International Baccalaureate Diploma Program (IBDP) in Grade 11 and 12, as mandated by the school bylaws and regulations. The sample was a mono-cultural group as all students were Turkish and come mainly from eastern Turkey. All students studied their middle schools (mainly at government state schools) with the national curriculum. All students spoke Turkish as a native tongue and spoke English as a foreign language (mostly at advanced level). The students had diverse socio-economic status from privileged to underprivileged at the case school. All students were on either partial or full academic merit scholarships and some underprivileged students were getting financial support (additional scholarships for meals, books, service bus, etc.) from the school. All the sample group entered into the school with a two-stage admission examination which consisted of components in languages, sciences, math, nonverbal reasoning as well as music and art components. The entire sample group took the IGCSE exams at the end of Grade 10 when they are aged 15-16. At the end of the two-year IGCSE program, students were admitted to the IBDP with the same selection criteria. All the sample

groups who were only successful in the IGCSE examinations took IBDP exams at the end of Grade 12 when they were aged 17-18. At the end of the IBDP program, the entire sample group got their national high school diploma and graduated from this case school when they were granted the IB diploma.

2.3. Data Collection

The data consisted of internal assessment marks and external assessment marks of the IBDP students (*N*=338, 164 male and 174 female) who were enrolled at the case school between the years of 2012 to 2019. The year 2020 was not included since the IB announced that the May 2020 final exam series would not take place due to the Covid (19) coronavirus pandemic. Specifically, the data obtained from the years of 2012 to 2019 consisted of the sample group's internal assessment scores of Turkish, English, mathematics, physics, chemistry, and biology courses at the standard level (SL) and higher level (HL) as well as the external assessment scores of Turkish, English, mathematics, physics, chemistry, and biology courses at the standard level (SL) and higher level (HL).

Internal assessment scores collected for data analysis were based upon the marks given by the teachers for students' project work, fieldwork, laboratory practical work, mathematical investigations, essays, or oral commentaries in the IBDP courses. The independent variable of the study is the final internal assessment score of each course. These marks were externally moderated by the IB examiners. The dependent variable of the study is the IBDP final external assessment score of each course. External assessment scores were based upon the marks given by the IB examiners (not the school's teachers) for students' examination papers of the IBDP subjects set by the IB at the end of their study in the IBDP. The final external assessment score is based upon the internal assessment score and external assessment score, but external and internal assessment components have different weightage in awarding of the final marks in the IBDP. In science subjects (physics, chemistry and biology), internal assessment contributes 20% to the final mark; whereas, external assessment contributes 80% to the final mark. As for the languages (Turkish Literature and English Language and Literature), internal assessment contributes 30% to the final mark and external assessment contributes 70% to the final mark. In mathematics, the contribution of internal assessment is 20% and the contribution of external assessment is 80% to the final mark.

The IBDP students are required to take external examinations in six different courses (three at the standard level and three at the higher level), which they select based on their career and occupational choices. Therefore, the IBDP total outcome score for six IBDP courses is 42. However, as a holistic international education program, the IBDP also includes Creativity, Activity and Service, Theory of Knowledge and Extended Essay core components, from which students may receive up to three bonus points, which make the final IBDP total outcome score as 45. The marks are officially announced to the IB schools on the International Baccalaureate Information Systems (IBIS) by the International Baccalaureate Organization. These data were retrieved electronically from the case school's IBIS account in its original format and collected with the permission from the school. The school administration gave an ethical approval to undertake and publish this current research.

2.4. Data Analysis

This study utilized quantitative methods to investigate the relationship between internal and external assessment in the IBDP context. In this study, the independent variables (predictors) are internal

assessment scores of IBDP SL and HL Turkish, English, mathematics, physics, chemistry, and biology courses. Dependent variables are the scores of external assessment scores of IBDP SL and HL Turkish, English, mathematics, physics, chemistry, and biology courses. Descriptive statistics were obtained first through SPSS for all independent and dependent variables. Then, Pearson correlation coefficient values were examined in order to see the relationship between the internal assessment scores and external assessment scores in the pertinent aforementioned IBDP courses. Finally, the data were analysed with simple linear regression in order to see to what extent the internal assessment scores explain the variation in the external assessment scores in IBDP.

Statistical assumptions were checked before conducting a linear regression for prediction of dependent variables, higher level and standard level externally assessed components in Biology, Chemistry, English, Mathematics, Physics and Turkish, with internal assessment components of higher level and standard level subjects in Biology, Chemistry, English, Mathematics, Physics and Turkish. The first assumption is normality. In order to check for normality, visual inspection of histogram, examination of Skewness and Kurtosis values have been conducted. Skewness values ranged from -0.70 to 0.64 while Kurtosis values ranged from -0.32 to 0.93. Analysis of Kolmogorov-Smirnov test and Skewness / Kurtosis values provide us with information that the data shows 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). We can assume that homoscedasticity was satisfied for the data. If we let external assessment score is the outcome, and internal assessment score a predictor (or explanatory variable), then the below equation is used to express the structural model: $E(Y|x) = \beta_0 + \beta_1 x$. If we put our variables into this equation, then the model can be presented as: EA Score (y) = $\beta_0 + \beta_1$*(IA Score). In this equation, β_0 represents the intercept (seen as *constant* in the unstandardized coefficients section in Table 7 and 8) and β_1 stands for slope (gradient) coefficient. This model shows that if the slope coefficient value is not zero ($p < 0.05$) for a particular course, a change in IA score has an effect on EA score.

3. FINDINGS

3.1. Descriptive Statistical Results

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.

Table 1 below shows the descriptive statistics for the higher level external assessment components in biology, chemistry, English, math, physics and Turkish.

Table 1: *Descriptive Statistics of External Assessment of High Level Subjects*

Descriptive statistics for HL Subjects External Assessment						
	Biology	Chemistry	English	Math	Physics	Turkish
N	176	206	111	132	109	194
Mean	46.52	42.77	44.69	45.76	36.89	52.92
Median	47.46	43.30	43.80	45.46	35.91	53.00
Mode	45.50	45.40	34.75	38.60	11.75	54.0
Std. Deviation	9.21	9.68	8.93	10.31	10.24	6.92
Skewness	-.29	-.007	.31	.12	.37	-.33
Std. Error of Skewness	.18	.17	.23	.21	.23	.17
Kurtosis	.13	-.09	-.63	-.49	-.29	.28
Std. Error of Kurtosis	.36	.33	.45	.42	.46	.35
Range	52.30	52.45	39.75	47.75	50.29	39.0
Minimum	18.50	13.90	25.25	25.58	11.75	30.0
Maximum	70.80	66.35	65.00	73.33	62.03	69.0

Table 2 below shows the descriptive statistics for the higher level internal assessment components in biology, chemistry, English, math, physics and Turkish.

Table 2: *Descriptive Statistics of Internal Assessment of High Level Subjects*

Descriptive statistics for HL Subjects Internal Assessment						
	Biology	Chemistry	English	Math	Physics	Turkish
N	176	206	111	132	109	194
Mean	14.56	15.20	20.49	13.41	15.54	25.54
Median	14.50	15.0	20.00	13.00	16.00	25.0
Mode	16.50	15.00	19.0	13.00	15.00	24.0
Std. Deviation	3.03	3.02	3.11	2.15	2.06	2.02
Skewness	-.14	.026	.38	-.37	-.46	.29
Std. Error of Skewness	.18	.17	.23	.21	.23	.17
Kurtosis	.45	-.76	-.49	.38	-.31	-.64
Std. Error of Kurtosis	.36	.34	.45	.42	.46	.34
Range	18.00	13.67	14.0	11.00	9.00	9.0
Minimum	4.00	8.33	13.0	8.00	10.50	21.0
Maximum	22.00	22.00	27.0	19.00	19.50	30.0

Table 3 below shows the descriptive statistics for the standard level external assessment components in biology, chemistry, English, math, physics and Turkish.

Table 3: *Descriptive Statistics of External Assessment of SL Subjects*

Descriptive statistics for SL Subjects External Assessment						
	Biology	Chemistry	English	Math	Physics	Turkish
N	74	40	234	217	71	155
Mean	45.51	45.18	40.84	60.33	39.55	52.65
Median	45.99	45.50	39.00	62.16	39.52	54.00
Mode	36.15	22.65	45.25	58.67	19.07	55
Std. Deviation	8.42	9.57	9.44	11.62	10.01	6.77
Skewness	-.19	.04	.64	-.70	-.05	-.50
Std. Error of Skewness	.28	.37	.16	.16	.28	.19
Kurtosis	.93	-.14	-.132	-.06	-.70	-.08
Std. Error of Kurtosis	.55	.73	.32	.33	.56	.39
Range	47.80	42.21	45.31	52.00	42.97	35
Minimum	18.35	22.65	22.80	27.11	19.07	32
Maximum	66.15	64.86	68.11	79.11	62.04	67

Table 4 below shows the descriptive statistics for the standard level internal assessment components in biology, chemistry, English, math, physics and Turkish.

Table 4: *Descriptive Statistics for Internal Assessment of SL Subjects*

Descriptive statistics for SL Subjects Internal Assessment						
	Biology	Chemistry	English	Math	Physics	Turkish
N	74	40	234	217	71	155
Mean	12.72	14.67	22.41	12.08	15.53	26.18
Median	12.50	14.16	22.00	12.00	16.00	26.00
Mode	11.50	13.33	19.0	12.00	16.50	27.0
Std. Deviation	2.48	2.37	4.19	2.29	2.25	1.92
Skewness	.12	.39	.29	-.24	-.69	-.33
Std. Error of Skewness	.28	.37	.16	.16	.28	.19
Kurtosis	-.15	-.02	-.68	.31	.56	-.15
Std. Error of Kurtosis	.55	.73	.32	.33	.56	.39
Range	11.00	11.00	21.0	14.00	11.67	8.0
Minimum	7.00	10.00	11.0	6.00	8.33	22.0
Maximum	18.00	21.00	32.0	20.00	20.00	30.0

3.1.1. Simple Linear Regression Analysis Results

Simple linear regression model was used in order to see to what extent IBDP internal assessment scores predict the IBDP external assessment scores in higher level (HL) biology, chemistry, English, mathematics, physics and Turkish courses. As seen in Table 5 below, 11% (*R^2= 0.11; R2 adjusted =0.11*) of the variation in IBDP HL Biology external assessment component can be explained by the model including only IBDP HL Biology internal assessment scores. For DP HL Chemistry external assessment 16% (*R^2= 0.16; R2 adjusted =0.16*), for IBDP HL English external assessment 39% (*R^2= 0.39; R2 adjusted =0.39*), for IBDP HL Mathematics external assessment 11% (*R^2= 0.11; R2 adjusted =0.11*), for IBDP HL Physics external assessment 4% (*R^2= 0.04; R2 adjusted =0.04*), for IBDP HL Turkish external assessment 7% (*R^2= 0.07; R2 adjusted =0.07*) of the variations can be explained by this model.

Table 5: *Regression Adjusted R^2 Results for IBDP HL Courses*

	Biology EA	Chemistry EA	English EA	Mathematics EA	Physics EA	Turkish EA
Biology IA	.11					
Chemistry IA		.16				
English IA			.39			
Math IA				.11		
Physics IA					.04	
Turkish IA						.07

Simple linear regression model was also used in order to see to what extent IBDP internal assessment scores predict the IBDP external assessment scores in standard level (SL) biology, chemistry, English, mathematics, physics and Turkish courses. As seen in Table 6 below, 29% (*R^2= 0.29; R2 adjusted =0.29*) of the variation in IBDP SL Biology external assessment component can be explained by the model including only IBDP SL Biology internal assessment scores. For DP SL Chemistry external assessment 10% (*R^2= 0.10; R2 adjusted =0.10*), for IBDP SL English external assessment 0% (*R^2= 0.00; R2 adjusted =0.00*), for IBDP SL Mathematics external assessment 16% (*R^2= 0.16; R2 adjusted =0.16*),

for IBDP SL Physics external assessment 18% (R^2= *0.18; R2 adjusted =0.18*), for IBDP SL Turkish external assessment 7% (R^2= *0.07; R2 adjusted =0.07*) of the variations can be explained by this model.

Table 6: *Regression Adjusted R^2 Results For IBDP SL Courses*

	Biology EA	Chemistry EA	English EA	Mathematics EA	Physics EA	Turkish EA
Biology IA	.29					
Chemistry IA		.10				
English IA			.00			
Math IA				.16		
Physics IA					.18	
Turkish IA						.07

The 'B' column in the Table 7 below unstandardized coefficients section indicates the values of the gradient (β_1) and intercept (constant = β_0) terms for the regression line. The model is: IBDP external assessment score (y) = β_0+ β_1*(IBDP internal assessment score). The "β_1" value against IBDP internal assessment score shows slope coefficients. For all courses gradient values are above 0 and this can be tested by sig values against IBDP internal assessment scores. Sig values presented in Table 7 and Table 8 below show significant evidence that the gradient values are not 0 ($p < 0.05$). This shows that a change in IBDP internal assessment score has an effect on IBDP external assessment score. The slope coefficient for IBDP HL biology internal assessment score is 1.023 so that IBDP HL biology external assessment score increases by 1.023 for each 1 point increase in IBDP HL biology internal assessment score (β = 1.023; $p < 0.05$). IBDP HL chemistry external assessment score increases by 1.291 for each 1 point increase in IBDP HL chemistry internal assessment score (β = 1.291; $p < 0.05$). IBDP HL English external assessment score increases by 1.822 for each 1 point increase in IBDP HL English internal assessment score (β = 1.822; $p < 0.05$). IBDP HL math external assessment score increases by 1.608 for each 1 point increase in IBDP HL math internal assessment score (β = 1.608; $p < 0.05$). IBDP HL physics external assessment score increases by 1.096 for each 1 point increase in IBDP HL physics internal assessment score (β = 1.096; $p < 0.05$). IBDP HL Turkish external assessment score increases by 0.946 for each 1 point increase in IBDP HL Turkish internal assessment score (β = 0.946; $p < 0.05$).

Table 7: *Unstandardized and Standardized Regression Coefficients for IBDP HL EA Components*

Model	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t.	Sig.
Constant	31.625	3.223	.337	9.812	.000
IBDP Biology IA	1.023	.217		4.721	.000
Constant	23.150	3.181	.403	7.277	.000
IBDP Chemistry IA	1.291	.205		6.289	.000
Constant	7.569	4.422	.631	1.712	.090
IBDP English IA	1.822	.213		8.491	.000
Constant	24.192	5.372	.336	4.503	.000
IBDP Math IA	1.608	.395		4.067	.000

Constant	19.861	7.347	.221	2.703	.008
IBDP Physics IA	1.096	.469		2.339	.021
Constant	28.768	6.076	.277	4.735	.000
IBDP Turkish IA	.946	.237		3.987	.000

As indicated in Table 8 below, the slope coefficient for IBDP SL biology internal assessment score is 1.840, so IBDP SL Biology external assessment score increases by 1.840 for each 1 point increase in IBDP SL biology internal assessment score ($\beta = 1.840$; $p < 0.05$). IBDP SL chemistry external assessment score increases by 1.432 for each 1 point increase in IBDP SL chemistry internal assessment score ($\beta = 1.432$; $p < 0.05$). IBDP SL English external assessment score increases by -0.139 for each 1 point increase in IBDP SL English internal assessment score ($\beta = -0.139$; $p < 0.05$). IBDP SL math external assessment score increases by 2.056 for each 1 point increase in IBDP SL math internal assessment score ($\beta = 2.056$; $p < 0.05$). IBDP SL physics external assessment score increases by 1.959 for each 1 point increase in IBDP SL physics internal assessment score ($\beta = 1.959$; $p < 0.05$). IBDP SL Turkish external assessment score increases by 0.979 for each 1 point increase in IBDP SL physics internal assessment score ($\beta = 0.979$; $p < 0.05$).

Table 8: *Unstandardized and Standardized Regression Coefficients for IBDP SL EA Components*

	Unstandardized Coefficients		Standardized Coefficients		
Model	B	Std. Error	Beta	t.	Sig.
Constant	22.109	4.339	.544	5.095	.000
IBDP Biology IA	1.840	.335		5.495	.000
Constant	24.174	9.078	.355	2.663	.011
IBDP Chemistry IA	1.432	.611		2.344	.024
Constant	43.953	3.367	-.062	13.055	.000
IBDP English IA	-.139	.148		-.940	.348
Constant	35.499	3.886	.406	9.135	.000
IBDP Math IA	2.056	.316		6.506	.000
Constant	9.111	7.539	.441	1.209	.231
IBDP Physics IA	1.959	.480		4.081	.000
Constant	27.032	7.182	.278	3.764	.000
IBDP Turkish IA	.979	.274		3.577	.000

3.1.2. Correlation Analysis Results

The study found that there is a moderate relationship between internal assessment and external assessment in IBDP HL English. Also, Pearson correlation analysis showed a weak relationship between internal assessment and external assessment in IBDP HL biology, chemistry, mathematics, physics and Turkish courses, as shown in Table 9 below.

Table 9: *Correlations for Internal Assessment and External Assessment in IBDP HL Courses*

	Biology EA	Chemistry EA	English EA	Mathematics EA	Physics EA	Turkish EA
Biology IA	.33					
Chemistry IA		.40				
English IA			.63			
Math IA				.33		
Physics IA					.22	
Turkish IA						.28

As shown in Table 10 below, Pearson correlation analysis for the IBDP SL courses show that there is a moderate relationship between internal assessment and external assessment in IBDP SL Biology course. Also, Pearson correlation analysis showed a weak relationship between internal assessment and external assessment in IBDP SL chemistry, English, mathematics, physics and Turkish courses.

Table 10: *Correlations for Internal Assessment and External Assessment in IBDP SL Courses*

	Biology EA	Chemistry EA	English EA	Mathematics EA	Physics EA	Turkish EA
Biology IA	.54					
Chemistry IA		.35				
English IA			.06			
Math IA				.40		
Physics IA					.44	
Turkish IA						.28

4. DISCUSSION and RESULT

The aim of the current study was to explain the relationship between student performance in internal assessment (IA) and external assessment (EA) in the IBDP courses as well as to explore the predictors of the IBDP external assessment scores in languages (Turkish and English), sciences (physics, chemistry, biology), and mathematics. Specifically, in terms of languages (Turkish and English), the current study found that there is a moderate correlation only between IBDP HL English IA and EA components, and 39% of the variation in IBDP HL English EA scores can be explained by the model containing only IBDP IA scores, but only 7% of the variation in IBDP SL English EA scores can be explained by the same model. It should be noted that SL IA mean scores are higher than HL IA mean scores and interestingly HL EA mean scores are higher than SL EA mean scores. At this point, it should be noted that students who are strong in English prefer studying HL English, yet their IA mean score is lower than SL students. That said, their EA mean score is still higher than SL students, so there may be some other factors explaining SL students' success in their IA. When it comes to the Turkish, the study indicated that the extent of prediction by Turkish IA scores for Turkish EA scores is very small (both for SL and HL only 7%). This can be explained by considering students' increasing intrinsic motivation on working on the topic of their choice for their IA, and also in the case study school, almost all students got high scores from Turkish IA and Turkish EA, and so there was already small variation in these scores.

As for mathematics and sciences, the results revealed that a small percentage – only 16% (in SL) and 11% (in HL) for mathematics and 18% (in SL) and 4% (in HL) for physics – of the variation in IBDP EA scores can be explained by the model containing only IA scores. Thus, some other factors – such as

the teacher input, exam anxiety and time pressure, experience of teachers, motivation – may have a critical role in affecting students' IBDP EA scores (Thompson and Correa, 1989). Additionally, the study also indicated a moderate relationship between the IA scores and the EA scores for SL biology 54%, but in HL biology and SL and HL chemistry, the relationship between the IA scores and the EA scores are weak (33% in HL biology, 40% in HL chemistry, and 35% in SL chemistry). For both biology and chemistry, only small percentages of the variations in IBDP EA scores can be explained by the model containing only IBDP IA scores (29% in SL biology, 11% in HL biology, 16% in HL chemistry, and 10% in SL chemistry).

To what extent the syllabus covered during DP years are related to the internal assessment done by students may be an important factor to explain the degree of strength of internal assessment scores as a predictor on external assessment scores for a particular course. For example, in science courses or mathematics, a student may choose a research topic which is beyond the syllabus and write a good report through extensive literature research and good academic writing skills. As a result, students may get a remarkably high grade from their internal assessment component. At this point, it is difficult to claim that the students have a satisfactory or more than satisfactory knowledge of the science or mathematics IBDP syllabus by taking only IA grade into account because its content is out of the syllabus.

EA performances only depend upon the individual capabilities of students as these scores are earned as a result of written exams (in certain time allocations) administered under supervision of teachers. But, for internal assessment studies, students may get outside help easily from professionals as these works are completed in a large scope of time – more than 1 year – in which a student may submit an excellent draft which does not reflect his/her real skills and capabilities. It is reported in the literature that it is difficult to separate a student's academic achievement from his/her study-homework behaviours, but it is also reported that especially in high school level students get more parental support -in different ways- for their homework (Núñez, et al., 2015; Gonida, 2014). Internal assessment works in the IBDP program can be regarded also as homework because students complete their works mostly out of school without teacher supervision. In this regard, it is exceedingly difficult to measure the level of support students get to complete their work which is expected to reflect their own academic performance. The earlier studies conducted by Chansarkar, (1987), Elton (1998) and Simonite (2003) indicate that increased use of academic coursework in assessment of students yields higher overall marks but smaller variation between students which affects quality of the assessment in a negative way.

Contrary to the findings of a study (conducted with a sample of 25,875 high school students) which claimed that "academic coursework had direct effects on achievement" (Keith and Cool, 1992, p. 207), the current study indicated that even submitting excellent coursework does not mean that students will have high scores from exams. The completion of an academic coursework has its own conditions and focuses on different abilities and skills of students, such as research, creativity, communication, collaboration and so forth. However, exams have different conditions in which a student can use only his/her own knowledge in a limited time, so along with some of aforementioned skills and abilities, exams aim to test content knowledge, thinking skills, reading and comprehension skills, scientific literacy and scientific reasoning skills, and time management skills (Elton, 1988).

In order to explain the reasons why internal assessment works are not strong predictors of the scores obtained from external assessment, examination anxiety and time pressure should be taken into consideration, as well. These two factors may affect students' mental and physical well-being and prevent them from showing their real academic performance. On the other hand, completing internal

assessment works over a long-time span, students have no time pressure and can express themselves in a more clear and comfortable way which helps to exhibit their real level of skills and abilities. Thus, students usually want to be assessed by only academic coursework or a mixture of academic coursework and examinations which yield higher grades for them (Richardson, 2015). Another point which should be considered is that internal assessment works of students are internally assessed and externally moderated. It is possible that teachers may give higher marks if they are held accountable to students or institutions for low marks they give (Bridges et al., 2002). These internally assessed works are moderated by external examiners, but earlier research in the literature indicates that it is difficult to reach a high level of standardization between examiners, and procedural documentation's contribution is limited to the reliability of the assessment (Elton, 1998; Yorke, et al., 2000).

When the results and findings of the current study are taken into account, it can be put forward that a high mark from internal assessment work of a particular course does not indicate that students will perform in the same way in external examinations of the same course. Earlier studies show that the inclusion of academic coursework can make overall assessment stronger, but its weight should be carefully determined, and examinations should be used as a final assessment tool (Gibney, 2013). Academic coursework should not create an unfair advantage to an academically low achieving student (Richardson, 2015; Simonite, 2003). IBDP grades are especially important for students as they may use these grades for university and scholarship applications, and to obtain higher grades from their academic coursework they may exhibit unacceptable behaviours such as plagiarism, collusion or contract cheating, so policy makers and educators should not overlook this reality when they make their decisions on the assessment models (Richardson, 2015; Bridges et al., 2002).

This current study concludes that IA scores explain a small percentage of the variation in IBDP examination scores, except for English HL course. Hence, other factors such as study habits, time management skills, teacher support, exam anxiety and time pressure, parental involvement may have important roles to explain the variations in the IBDP examination scores. The results of this study indicate that increasing internal assessment works' weight in overall assessment may decrease the reliability of final grading. For challenging and unprecedented times such as Covid-19, institutions and policy makers should make different plans and use different systems for a fair assessment of students. This may include recording predicted grades at the end of each semester, collecting and recording data from in school examinations throughout the IBDP years, as well as creating reliable and accessible online examination systems.

This study acknowledges the following limitations within its scope. First of all, this research has only collected the internal and external assessment data from only one particular school context, so replicated studies may include more extensive data from different school contexts with a larger sample size. Additionally, the current research only focused on Turkish, English, mathematics, physics, chemistry and biology courses at the standard and high level for exploring the relationship between internal and external assessment, so future replicated studies may also focus on other subjects such as Language B, Economics, Visual Arts or other IB approved school-based courses (e.g., Turkey in the 20th century - TITC).

Finally, the authors of this study recommend some other further research with regard to the analysis of assessment in the IBDP context. The relationship between the teachers' judgments of student achievement (given in the form of predicted grades) and external assessment scores can be researched to

explore to what extent the overall teachers' estimated grades for different subjects predict students' overall performance in the external examinations in the IBDP. The findings of this future research may also yield some important implications for educators in terms of developing alternative assessment models in the final award of student achievement marks on their diplomas.

Declaration of Contribution Rate of Researchers

Author 3 contributed to collection of data and analysis of data, Author 1 contributed to introduction, literature review and method sescion of paper. Author 2 contributed discussion and conclusion section of the study. All authors provided feedback on all sections of the study.

Conflict of Interest Declaration

The authors declare that there is no conflict of interest.

REFERENCES

Awomolo, A. (1992, November). *The Challenges of combining internal and external assessment in certificate examinations: The West African examinations council experience*. [Paper presentation]. Annual Conference of the American Evaluation Association, Seattle, WA, USA.

Barrance, R. (2019). The fairness of internal assessment in the GCSE: The value of students' accounts. *Assessment in Education: Principles, Policy & Practice*, *26*(5), 563-583. Doi: 10.1080/0969594X.2019.1619514.

Badyal, D.K., & Sharma, M. (2020). Internal assessment in new MBBS curriculum: Methods and logistics. *International Journal of Applied Basic Medical Research*, *10*(2), 68-75.

Benton, T. (2016). Evidence for the reliability of coursework. In G.A. Nortvedt (Ed.), *17th Annual association of educational assessment, social and political underpinnings of educational assessment: Past, present and future conference proceedings* (1-173). Limassol: University of Cyprus.

Bhadwal, S. C., & Kishor, V. (2012). Relationship between internal and external assessment scores in Bachelor of Education Programme of Himachal Pradesh University. *Scholarly Research Journal for Interdisciplinary Studies*. *1*(3), 621-634.

Bridges, P., Cooper, A., Evanson, P., Haines, C., Jenkins, D., Scurry, D., Woolf, H., & Yorke, M. (2002). Coursework marks high, examination marks low: Discuss. *Assessment & Evaluation in Higher Education*, *27*(1), 35-48. Doi: 10.1080/02602930120105045.

Brown, R. (2002). Cultural dimensions of national and international educational assessment in Hayden, M, Thompson, J and Walker, G (eds). *International education in practice: dimensions for national and international schools*. London: Kogan Page.

Bullock, K., Bishop, K., Martin, S., & Reid, A. (2002). Learning from coursework in English and geography. *Cambridge Journal of Education*, *32*(3), 325–340.

Chansarkar, B. A., & Raut-Roy, U. (1987). Student performance under different assessment situations. *Assessment and evaluation in Higher Education*, *12*(2), 115-122.

Crisp, V., & Green, S. (2013). Teacher views on the effects of the change from coursework to controlled assessment in GCSEs. *Educational Research and Evaluation*, *19*(8), 680-699. Doi: 10.1080/13803611.2013.840244.

Ekolu, S. O. (2016). Correlation between formative and summative assessment results in engineering studies. In A. J. Swart (Ed.), *The 6th African engineering education association conference proceedings* (1-211). Bloemfontein: Central University of Technology.

Elton, L. (1998). Are UK degree standards going up, down or sideways? *Studies in Higher Education*, *23*(1), 35-42.

Elton, L. (1988). Student motivation and achievement, *Studies in Higher Education*, *13*(2), 215-222.

Elwood, J. (1999). Equity issues in performance assessment: The contribution of teacher-assessed coursework to gender-related differences in examination performance. *Educational Research and Evaluation*, *5*(4), 321–344.

Gibney, E. (2013). "Tailor work to cut out essays to order." Times Higher Education, June 13, 10. http://www.timeshighereducation.co.uk/news/bournemouth-builds-armoury-againstcontract- essay-cheats/2004757.article.

Gonida, E. N., & Cortina, K. S. (2014). Parental involvement in homework: Relations with parent and student achievement-related motivational beliefs and achievement. *British Journal of Educational Psychology*, *84*(3), 376-396.

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

International Baccalaureate (IB). (n.d.) IB facts and figures. Retrieved from https://www.ibo.org/about-the-ib/facts-and-figures/.

Jagesic, S., & Wyatt, J. (2018). *Postsecondary course performance of ap exam takers in subsequent coursework: Biology, languages, and studio art*. New York: College Board.

Kandamby, G. W.T.C. (2017). Coursework for promoting teaching and learning. *International Journal of Higher Education*, *6*(6), 129-138.

Keith, T. Z., & Cool, V. A. (1992). Testing models of school learning: Effects of quality of instruction, motivation, academic coursework, and homework on academic achievement. *School Psychology Quarterly*, *7*(3), 207-226.

Linn, M.C. (1992). Gender differences in educational achievement. In J. Pfleiderer (ed). *Sex equity in educational opportunity, achievement, and testing.* Princeton: Educational Testing Service.

Murdan, S. (2005). Exploring relationships between coursework and examination marks: A study from one school of pharmacy. *Pharmacy Education, 5*(2), 97–104. doi: 10.1080/15602210500140673.

Núñez, J. C., Suárez, N., Rosário, P., Vallejo, G., Valle, A., & Epstein, J. L. (2015). Relationships between perceived parental involvement in homework, student homework behaviors, and academic achievement: differences among elementary, junior high, and high school students. *Metacognition and Learning*, *10*(3), 375-406.

Richardson, J. T. E. (2015). Coursework versus examinations in end-of-module assessment: a literature review. *Assessment & Evaluation in Higher Education*, *40*(3), 439–455. http://dx.doi.org/10.1080/02602938.2014.919628.

Ross, J. A, & Kostuch, L. (2011). Consistency of report card grades and external assessments in a Canadian province. *Educational Assessment, Evaluation & Accountability*. *23*(2), 159-180. doi 10.1007/s11092-011-9117-3.

Scott, D. (1991). Issues and themes: Coursework and coursework assessment in the GCSE. *Research Papers in Education*, *6*(1), 3-19. doi: 10.1080/0267152910060102.

Simonite, V. (2003). The impact of coursework on degree classifications and the performance of individual students. *Assessment & Evaluation in Higher Education*, *28*(5), 459–470.

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

Tattersall, K. (1994). The role and functions of public examinations. *Assessment in Education:Principles, Policy & Practice*, *1*(3), 293–304.

Thompson, J. A., & Correa, L. H. (1989). A study of school and teacher inputs on student achievement outputs. *Journal of Education Finance*, *14*(3), 390-406.

Torrance, H. (2018). The return to final paper examining in English national curriculum assessment and school examinations: Issues of validity, accountability and politics. *British Journal of Educational Studies*, *66*(1), 3-27. Doi: 10.1080/00071005.2017.1322683.

Vitello, S., & Williamson, J. (2017). Internal versus external assessment in vocational qualifications: A commentary on the government's reforms in England. *London Review of Education, 15(*3), 536-548.

World health organization (WHO). (2020). *WHO director general's opening remarks at the media briefing on COVID*. Retrieved from https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---7-september-2020.

Wyatt, J., Jagesic, S., & Godfrey, K. (2018). *Postsecondary course performance of AP exam takers in subsequent coursework.* New York: College Board.

Yorke, M., P. Bridges, and H. Woolf. (2000). Mark distributions and marking practices in UK higher education: Some challenging issues. *Active Learning in Higher Education*, *1*(1), 7–27.

Ziderman, A. (1984). External examination versus internal assessment: The Israeli maturity certificate. *Comparative Education Review*, *28*(3), 477-484.