KURAM VE UYGULAMADA EĞİTİM BİLİMLERİ EDUCATIONAL SCIENCES: THEORY & PRACTICE

Received: September 24, 2016
Revisio"n received: January 25, 2017
Accepted: February 14, 2017
OnlineFirst: April 25, 2017

www.estp.com.tr
DOI 10.12738/estp.2017.4.0490 • August 2017 • 17(4) • 1255–1290

Research Article

Comparing the High School English Curriculum in Turkey through Multi-Analysis*

Veli Batdı[1]
Kilis 7 Aralık University

Abstract

This study aimed to compare the High School English Curriculum (HSEC) in accordance with Stufflebeam's context, input, process and product (CIPP) model through multi-analysis. The research includes both quantitative and qualitative aspects. A descriptive analysis was operated through Rasch Measurement Model; SPSS program for the quantitative aspect, and a thematic analysis was run as a case study for the qualitative aspect. The study includes the views of 95 out of 135 English teachers working at different types of Anatolian High Schools in Elazığ during 2014 and 2015 academic years. Twenty-seven teachers were selected for Rasch; 64 for SPSS and 26 for thematic analysis. The scale applied in Aközbek's study was used as a quantitative data collection tool. An interview form was used to collect qualitative data and analyses were performed through Maxqda-11 program. The reliability of data encoders was ensured. Rasch and SPSS results indicated that the 9th grade HSEC had the highest score, whilst the 12th had the lowest. The fact that there were more positive opinions for the 9th grade HSEC, whilst there were negative ones for the 12th grade; means the results of three analyses are consistent. Multi-analysis is proposed for use in studies in the future.

Keywords

English Curriculum • CIPP model • Multi-analysis • Curriculum evaluation • Rasch model • Maxqda program

* This study was presented as an oral paper at the Congress of 24th National Education Sciences held April 16th-19th, 2015 at Niğde University, Niğde Turkey.

1 **Correspondence to:** Veli Batdı (PhD), Department of Curriculum and Instruction, Muallim Rıfat Education Faculty, 7 Aralık University, Kilis 79000 Turkey. Email: veb_27@hotmail.com

Citation: Batdı, V. (2017). Comparing the high school English curriculum in Turkey through multi-analysis. *Educational Sciences: Theory & Practice, 17*, 1255–1290. http://dx.doi.org/10.12738/estp.2017.4.0490

It is a requirement that education systems are compatible with innovations in terms of technological, sociological, physiological and cultural changes and developments. Seeking quality in education has always been at the forefront of countries' agendas and the necessity and importance of this issue is inarguable. In this regard, curriculums have to be prepared systematically, and coordinated and evaluated regularly, in order to ensure the increase in the quality of education (Ball, 2011; Cook, 2010). The curriculum plays a significant role in disseminating and fulfilling the long and short term goals set by countries seeking to develop unity and solidarity. Furthermore, curriculums can be thought of as a means to develop individuals from social, political and economic points of view (Varış, 1998). Therefore, they have an important role in the development of both individuals and countries.

The examinations undertaken to investigate the effectiveness of curriculums with regard to planning, purpose, process and outcomes are defined as a means of curriculum evaluation (Setyoko, Tunas, & Sunaryo, 2016). Every phase of the curriculum should be conducted systematically in order to fulfill the above-mentioned functions effectively. In other words, curriculum evaluation (Demirel, 2009), review and development (Marsh & Willis, 2007) have to be conducted in order to make decisions about the effectiveness of the curriculum. Hence, curriculum evaluation enables the revision, development and continuity of the curriculum (Klenowski, 2010) and determines the effectiveness, success, functions and level of the educational attainment goals of the curriculum (Wolf, Hill, & Evers, 2006). Curriculum evaluation also determines whether or not the education provided is compatible with individuals' interests and needs, technological progress, changes and developments (Kali & Linn, 2010). In this respect, a number of models are used in the evaluation of curriculums, including Stake's Congruence-Contingency Model, the Kirkpatrick Model and Stufflebeam's Context, Input, Process and Product Model (CIPP). The current study aims to evaluate curriculums according to Stufflebeam's CIPP model.

The CIPP model is used as an evaluation model not only in the educational field, but also in various fields outside the education context, such as in the health field (Steinert, Cruess, Cruess, & Snell, 2005). The CIPP model covers all the phases of the curriculum in which educational planners and evaluators undertake a process of revision. It can be used as a model from the very beginning, even in the curriculum development phase, and provides a useful contribution by giving particular importance to the complicated content which has to be taken into account in curriculum evaluation (Gandomkar, Jalili, & Mirzazadeh, 2015). The model was first proposed by Stufflebeam and his colleagues in the early 1970s. It aims to arrive at a decision through current knowledge by making distinctions between the evaluative and managing aspects of tasks. The responsibility of the evaluative aspect is to determine, provide and present knowledge for those managers who are responsible

for achieving access to knowledge for arriving at correct decisions and the application of the evaluation results (Smith & Benjamaporn, 2013). The model consists of four basic processes. Context, which is the first phase, involves making systematic changes in the current curriculum or making it appropriate for changing situations. The input phase determines the appropriate resources to meet the determined needs, and the process phase provides decisive knowledge for revisions and arrangements to be implemented in the planned curriculum. The last phase, product, generates important knowledge to justify the curriculum outputs (Stufflebeam & Shinkfield, 2007, as cited in Mirzazadeh et al., 2016). Every phase of the CIPP model applies to a different situation. The evaluations conducted through the model are comprehensive and detailed methods of curriculum evaluation and development. Therefore, an evaluation program which presents a comprehensive examination with regard to the needs, planning, aim, content, process and evaluation phases of a curriculum is more likely to contribute to the development of better quality curriculums.

Research has shown that curriculums which facilitate individuals' access to education have to be evaluated constantly and regularly to keep up-to-date, and to comply with changes and developments in the current conditions in which quality of education plays a significant role. In this regard, the curriculum evaluation models have to be taken into account when making decisions in relation to the future of the curriculum (Fitzpatrick, Sanders, & Worthen, 2004). At this point, objective and realistic evaluations can be undertaken by making use of Stufflebeam's CIPP Model. For this purpose, it has been stressed that curriculums must be evaluated in terms of utility, feasibility, propriety and accuracy standards (Stufflebeam, 1994). Although curriculum evaluation studies, which are quite important for education, have been encountered at national and international levels (Clewell et al., 2004; Duru & Korkmaz, 2010; Orbeyi & Güven, 2008; Rolstad, Mahoney, & Glass, 2005; Şahin, 2007), a different approach has been embraced in the current research utilizing analyses such as the Rasch model, SPSS and Maxqda programs to emphasize the necessity of curriculum evaluation and draw attention to this issue. It is considered that the results which were obtained from the evaluation of the High School English Curriculum (HSEC) in the context of the CIPP model will make a significant contribution to the literature.

The Aim and Significance of the Research

This research aimed to interpret and compare the findings from the evaluation of the HSEC with respect to the context, input, process and product dimensions of the CIPP model through the Rasch, SPSS and Maxqda programs. A variety of studies (Arıbaş & Tok, 2004; Gökdemir, 2005; Işık, 2008) have indicated that teaching English, generally, has not been able to achieve the success that was aimed for in

Turkey. It can be argued that one of the reasons for this is that the evaluation of the curriculum at regular and systematic periods and the necessary regulations are not conducted correctly. Therefore, the current research investigated this problem and considered suggestions for a solution by examining the HSEC based on the CIPP model. However, when the studies concerned with the evaluation of the English teaching curriculum are examined, it is apparent that those evaluations do not involve multidimensional analysis. In other words, it was apparent that multi-analysis was not undertaken in those evaluations using different programs and comparisons (Ersen Yanık, 2007; Parmaksız & İncirci, 2016; Zehir Topkaya & Küçük, 2010).

The present research study aimed to evaluate the HSEC from different angles and compare these evaluations to each other in order to reach a common and general result. Therefore, this research is expected to attract attention in the scientific field as it involves differences. When the CIPP model was reviewed in the literature at both national and international levels (Brewer, 2010; Karataş & Fer, 2011; Tseng, Diez, Lou, Tsai, & Tsai, 2010) or in general terms in curriculum evaluation (Bauer, 2008; Creswell, 2002; Diamond, 1998; Mirzazadeh et al., 2016; Özdemir, 2009; Sağlam & Yüksel, 2007), a number of studies were encountered. However, it was confirmed that these studies generally used one dimensional data analysis (quantitative or qualitative), unlike the current study which used a multiple analysis. Thus, this study is expected to make a genuine contribution to the literature considering its extensive scope and multi-dimensional results in both general and specific senses.

Multiple analysis was conducted in this research. Multiple analysis aims to reach comprehensive and common results by evaluating the data obtained from different resources through different analyses. In this sense, the wealth of data resources is known to increase the quality of research by enabling research to be evaluated from different points of view (Marvasti, 2004; Patton, 2002). In other words, multiple analysis evaluates a current situation for any subject with different programs and presents complementary, supplementary and scientific knowledge to the analysis process. At this point, the study conducted by Cook (2010) was taken into account as a reference to use different analysis programs in this study. It emphasized the necessity to use different approaches in the curriculum evaluation. Furthermore, the purpose of using different analysis methods in multiple analysis is to complement insufficiencies, by comparing these analyses to one another, or to enable access to knowledge which cannot be obtained through a particular analysis type. Thus, the main aim of this analysis was to supplement the lack of knowledge by the means of different programs, make a detailed analysis and present all the knowledge about the related issues in clear and concise detail. In this context, the presentation of insufficient knowledge, which can be encountered in research, can be handled by using more than one analysis program.

Methodology

The present study consists of both quantitative and qualitative aspects. These are the Rasch measurement model, SPSS 17 and Maxqda 11 programs. The Rasch measurement model provides a framework for the review of rating scale structures. It can present helpful information for commenting on the meaning of ratings specified in large-scale performance evaluation contexts (Engelhard & Wind, 2013). The Rasch model also develops ways to measure such latent variables as human attributes including attitudes, opinions, interests, knowledge, traits, skills, proficiencies and aptitudes (Sjaastad, 2014). A Rasch model in which there are three facets, including levels of proficiency, items difficulty and severity/leniency of judges, necessitates the unification and standardization of these facets on a common plane in order to obtain objective and reliable measurements. In other words, this multi-faceted model allows the comparison of participants' performances, items difficulty and severity/leniency of judges simultaneously (Akın & Baştürk, 2012).

As stated earlier, the research aimed to obtain complementary information concerning a topic through different analytical programs in the multiple analysis process. The necessity of using Rasch, SPSS and thematic analyses in the present study is explained in detail in this section. The use of different analyses in a single study is expected to contribute to obtaining valid, reliable, and both general and specific information, based on data diversity. The Rasch model was used to conduct multi-product evaluations considering the outcomes of HSEC with respect to CIPP and to identify the participant bias. The SPSS program was performed to obtain comments on the participation levels of the respondents regarding the related products. Finally, the Maxqda program was applied to reveal participants' views on HSEC in the context of the CIPP model via visual models. The main purpose was to compare the findings reached following the analyses so as to make correlations. The general framework of objectives that the research intended to achieve through the multiple analysis process are as follows: (i) within the scope of the Rasch model, a general analysis of the views related to evaluation of HSEC in terms of CIPP, the analysis of the severity/leniency of judges (27), an item difficulty analysis to evaluate HSEC in terms of CIPP, and an analysis of any bias of judges. (ii) Within the scope of the Maxqda 11 program, the aim was to examine the views on HSEC regarding context, input, process and product. (iii) Within the scope of the SPSS 17 program, the aim was to evaluate teachers' views on HSEC in terms of the CIPP model considering gender and to compare the arithmetic mean of sub-dimensions in the scale.

Research Type and Design

In the quantitative dimension of the study, the Rasch Measurement Model and the SPSS program were used. The Rasch measurement model allows the prediction of

the response of an individual's performance to a particular property, and provides significant data for the development, change, and monitoring of current measures (Boone & Scantlebury, 2006). The quantitative dimension of the research was carried out in accordance with descriptive analysis. This is a technique involving the rearrangement, organization and processing of data to transform raw data into an easily understandable and interpretable form of descriptive information (Zikmund, 2003). Ensuring there is information about the current state, descriptive analysis (Karasar, 2012), also makes it easy to obtain sufficient information from the sample size (Büyüköztürk, Kılıç-Çakmak, Akgün, Karadeniz, & Demirel, 2009). Descriptive data were collected from teachers working in different school types in a cross-sectional manner. This type of research has the sufficiency to assess all of the variables at once, to cover a large sample with different properties, and to generalize the results to the community or population (Olsen & George, 2004).

In the qualitative dimension of the study, a thematic examination was conducted using the Maxqda program in accordance with the case study. A case study is a pragmatic method of using multiple data sources in which there is a wide-ranging in-depth event or situation in a real environment (Yin, 1994). Alternatively, it is a type of research providing detailed information for future research on the event or situation that has been put forward, and clearly indicates what needs to be taken into consideration. It is also known that in case studies, examining the effectiveness of a curriculum, a case study is intended to identify the success or failure of the curriculum. Therefore, the aim was to investigate the positive and negative aspects of HSEC with the related research type.

Study Participants

The study included the opinions of 95 of the 135 high school English teachers who worked in different types of high schools during the academic year of 2014 and 2015 in Elazığ, Turkey (Table 1), who were well-versed in the HSEC (9th, 10th, 11th, 12th grade). In order to collect qualitative data, 26 teachers (10 from the Anatolian High School, 5 from the Anatolian Vocational and Anatolian Imam Hatip High School and 11 from the Science and Anatolian Teacher Training High School) were consulted from among the above-mentioned teachers. In the Rasch analysis, the number of teachers whose views were considered was 27. While the teachers in the Rasch analysis were different from those in the qualitative aspect of the research, their numbers in terms of school types were the same, with the exception of one from the Anatolian High school group. The number of teachers reached in the context of the SPSS analysis in order to obtain quantitative data was chosen to be 64 from the teachers given in Table 1.

Table 1
Demographic Characteristics of the Participants

Variables		N	%
Gender	Female	63	66.32
	Male	32	33.68
Length of service (Year)	1-5	8	8.42
	6-10	26	27.37
	11-15	32	33.68
	16-20	22	23.16
	21 and Over	7	7.37
Type of School	Science High School	4	4.21
	Anatolian High School	55	57.89
	Anatolian Teacher Training High School	7	7.37
	Anatolian Technical and Vocational High School	29	30.53
Total		95	100

In the context of the Maxqda, Rasch and SPSS analyses, maximum variation sampling was considered in order to reach the related numbers of teachers. Maximum variation sampling allows for the identification of similar and different aspects of varied cases related to a problem, and probes a more comprehensive perspective of the related problem (Büyüköztürk et al., 2009). The total number of teachers (26, 27, 64) in the study was 95, due to the fact that the teachers working in the Science and Anatolian Teacher Training High Schools (n = 11) were selected for the related three analysis types as a common category in the study. On the other hand, when the demographic characteristics in Table 1 are examined, it is apparent that 66.32% of the total 95 teachers are female and 33.68% are male. Moreover, in the context of the length of service, the number of teachers having 6 to 20 years of service is 83%, which signifies that the participants who have given their opinions have sufficient experience. The last variable considered the type of school; this indicates that the number of teachers working at the Anatolian High school (close to 58%) is more than other school types, as the number of Anatolian High Schools in Elazığ is greater.

Data Collection

As a quantitative data collection tool (for the Rasch and SPSS analyses), the teacher form of a scale developed by Karataş and Fer (2009) and re-organized by Aközbek (2008) considering Stufflebeam's CIPP model was used. The five-point Likert type scale (Never: 1.00-1.80; Rarely: 1.81-2.60, Sometimes: 2.61-3.40, Usually: 3.41-4.20 and Always: 4.21-5.00) consisted of 27 items and 4 factors (context, input, process, product). It was re-arranged by the researcher in the light of the experts' views (one Professor, two Associate Professors, three Assistant Professors at the Department of Educational Sciences, one expert at the Department of Assessment and Evaluation, six English and two Turkish Philology teachers working at High Schools). The expressions related to "Mathematics" in the original form of the scale

were replaced with the expressions of "foreign language" according to the views of the experts. The final version of the scale, which had been re-arranged with a view to collecting teachers' opinions regarding the HSEC, included 25 items and its validity and reliability were ensured. While the Cronbach alpha internal consistency coefficients of factors were between the values of .66 and .85 in the original form, these values calculated for the current form of the scale were found to be between .68 and .90. The fact that the value of Cronbach alpha was close to + 1.00 means that the consistency and reliability of the scale was high (Özdamar, 2010). Mean, standard deviation, and t-test in binary comparisons were used in the SPSS analyses and the significance level of the study was regarded as .05.

The qualitative data of the research was collected by the researcher through an interview form developed in consideration of the literature review and expert opinions. The form consisted of open-ended questions, and teachers were asked to express their views on CIPP in terms of context, input, process and product. Qualitative data obtained from 26 teachers through the interview form was converted into codes under four themes. In this study, the fact that the researcher had been teaching English for a long time in Elazığ city centre and had been the school/region branch president facilitated reaching the sample group in different education regions and from different school types in this study.

Data Analysis

In studies comprised of more than one analysis type, all data obtained are used simultaneously or sequentially to establish a common unity (Creswell, Plano Clark, Gutmann, & Hanson, 2003). In the present study, all the data received were analyzed sequentially and the study was presented as a whole. Hence, the data obtained in the context of multiple analysis, where both quantitative (Rasch and SPSS) and qualitative (thematic) methods are used together (Bamberger, 2012), were analyzed separately, on the basis of presenting a general overview of the evaluation of HSEC in the context of CIPP.

From the quantitative analysis, the Rasch measurement model developed by Linacre (1993) was performed by the FACETS analysis program. According to the Rasch measurement model, the first of the three facets of the study consisted of 27 English judges, the second facet was the scale items (25 items) regarding CIPP, while the third facet was HSEC (9th, 10th, 11th and 12th grades). On the other hand, analyses related to arithmetic mean, standard deviation and factors of quantitative data were made through the SPSS program. Qualitative analysis was conducted by expressing the data related to the views as themes and codes in the context of the HSEC considering the CIPP model. Hence, the thematic analysis, which is a flexible and useful research tool that provides rich and detailed data, was used in this process

(Braun & Clarke, 2006). After the interview form was independently examined by the two data coders, the views expressed in terms of the main and sub goal of the study were recorded in the form of themes and codes. The themes created with titles of context, input, process and product were uploaded to the Maxqda program along with the relevant codes. Moreover, the Cohen Kappa agreement values for each theme were calculated separately to ensure the reliability of the research data coders. A value between .20 and below means weak; .21 and .40 below medium; .41 and .60 medium; .61 and .80 good; and .81 and 1.00 very good (Viera & Garrett, 2005). When the values related to the themes in the present study were calculated, .81 was achieved for context; .85 for input; .82 for process and .91 for product. Thus, it was concluded that these values were in very good agreement (Appendix 1). In addition, the teachers who had contributed to the qualitative dimension of the research were coded as $1P_M$ or $2P_F$ (a male participant numbered as 1 or a female participant numbered as 2) and these codes were used during the research when their citations were included.

Results

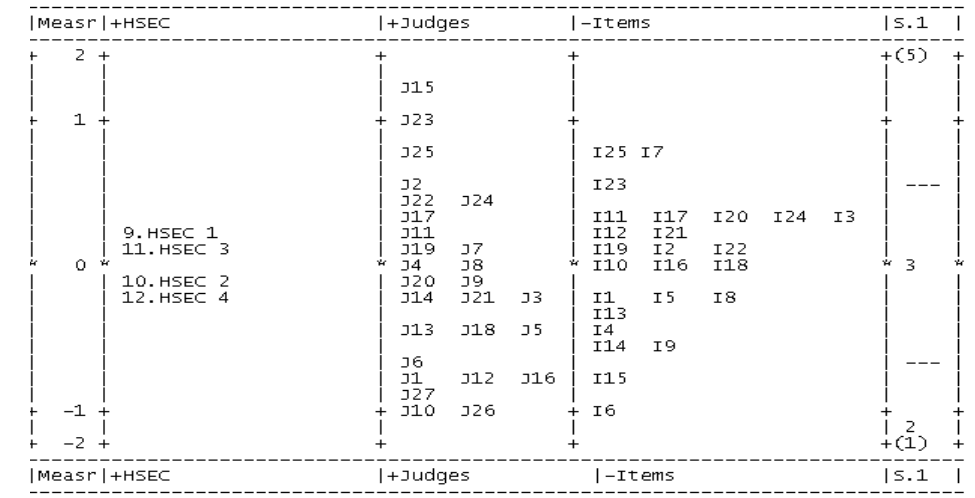
This section presents the quantitative and qualitative data obtained through multiple analysis in the research under separate headings. The tables for the Rasch and SPSS programs and the visual presentations of the Maxqda models are interpreted in detail. The findings, which are interpreted separately, are then combined in the conclusion section.

Quantitative Findings Related to Rasch Measurement Model and SPSS

Part of the findings obtained by the different analysis programs used in the research has been made accessible by the Rasch model. In this context, following the analysis of the English teachers' views about the evaluation of HSEC in terms of Stufflebeam's CIPP model via the multi-facetted Rasch measurement model. General information about these facets is presented in the calibration map in Figure 1.

The terms "HSEC" (9th, 10th, 11th, and 12th grades), "judges" and "criteria" were used to evaluate the curriculum according to the CIPP model and are indicated with different columns in the calibration map. The values between (-) and (+) on the left side of the map are the same logit measure for the facets. The qualities of the HSEC in the context of the CIPP model, severe/lenient scores of judges and the difficulty levels of the items are listed in the logit scale on this map. In the column related to HSEC, the curriculum coded as 9th HSEC has the highest score in the context of the CIPP model, while that coded as 12th HSEC has the lowest. When the data regarding judges' scores are considered, the judge coded J15 has the most lenient behavior, while J10 and J26 have the same and the most severe behavior.

It is apparent that the judges behaved differently in the scoring, but it is notable that there were also judges who exhibited equal scoring behaviors. When the criteria used for evaluating the HSEC are examined, it is apparent that M7 under the context factor and M25 under the product factor are the most difficult items, including the statement that "The course book can attract students' attention" and "The curriculum teaches language to students with pleasure." On the other hand, it is apparent that the item coded M6 in the input factors is the easiest item stating that "The project subjects are compatible with the course subjects."



HSEC: High school English Curriculum; J: Judge; I: item

Figure 1. Data calibration map.

In addition to the Rasch analyses, the research aimed to provide more detailed information by analyzing the level of arithmetic mean of the items on the basis of the grade variable via the SPSS Program. For this reason, when comparing the item arithmetic means of the Rasch measurement model and the results of the SPSS program in terms of grades, the lowest arithmetic mean in relation to the 9th ($\bar{x} = 2.41$), 10th ($\bar{x} = 2.35$), 11th ($\bar{x} = 2.41$)) and 12th grades ($\bar{x} = 2.55$) were in the I7 coded item in the SPSS analysis. The I6-coded item, on the other hand, had the highest value of the arithmetic mean in the context of the grade variable ($\bar{x}_{\text{9th grade}} = 3.84$, $\bar{x}_{\text{10th grade}} = 3.76$, $\bar{x}_{\text{11th grade}} = 3.62$, $\bar{x}_{\text{12th grade}} = 3.64$) (App. 2). Therefore, it is considered that the Rasch and SPSS analyses have consistent results with confirmatory and supportive findings. Additionally, when the arithmetic mean of the factors (context, input, process, and product) on the basis of grades were examined, the greatest mean value was in the 9th grade ($\bar{x}_{\text{Mean}} = \bar{x}_C = 3.44$, $\bar{x}_I = 3.04$, $\bar{x}_P = 3.33$, $\bar{x}_P = 2.85$) while the lowest was in the 12th grade ($\bar{x}_C = 2.97$, $\bar{x}_I = 2.88$, $\bar{x}_P = 2.95$, $\bar{x}_P = 2.60$). When the findings are

considered from the perspective of the CIPP model, it can be stated that the 9th grade curriculum has the highest score while the 12th grade has the lowest score.

Table 2

9th-12th HSEC Measurement Report

Obsvd Score	Obsvd Count	Obsvd Average	Fair Avrage	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	N	HSEC
2076	675	3.1	3.09	.17	.04	1.0	0	1.0	0	1	9.HSEC 1
2047	675	3.0	3.05	.12	.04	0.9	-1	0.9	-1	3	11.HSEC 3
1927	675	2.9	2.85	-.09	.04	0.9	-2	0.9	-2	2	10.HSEC 2
1854	673	2.8	2.74	-.21	.04	1.1	2	1.1	2	4	12.HSEC 4
1976.0	674.5	2.9	2.93	.00	.04	1.0	-0.1	1.0	-0.1	Mean (Count: 4)	
89.9	0.9	0.1	0.14	.15	.00	0.1	1.9	0.1	1.7	S.D.	

RMSE (Model) .04 Adj S.D. .15 Separation 3.55 Reliability .93
Fixed (all same) chi-square: 54.4 d.f.: 3 significance: .00
Random (normal) chi-square: 3.0 d.f.: 2 significance: .22

Table 2 presents a detailed measurement report covering the qualifications appropriate for the context, input, process and product stages of the CIPP (grades 9-12). The curricula are ranked according to the stages of the CIPP model considering the highest score to the lowest one. In this sense, the highest scored curriculum appeared to be the 9th HSEC and the lowest scored curriculum was the 12th HSEC. Furthermore, Table 2 shows that the standard error (RMSE-Root Mean Square Standard Error) for the logit values related to the qualities of the curricula was calculated to be .04. This value indicates the measurement error for all data except extreme values. The fact that the value was .04 can be interpreted as the standard error is quite low. The adjusted standard deviation value also appears to be below the critical value of 1.0 (Adj S.D. = .15) considering the value of error.

In the Rasch analysis, the reliability of HSEC in terms of the CIPP model is determined through the reliability coefficient. The reliability coefficient, which is determined as .93 via Rasch analysis, indicates that the programs are ranked with a very high level of confidence. In addition, when the other data in the table are examined in detail, and the chi-square test with the reliability index of 3.55 is performed with the reliability coefficient of .93, there is a statistically significant difference in terms of compliance with the context, input, process and product stages ($x^2 = 54.4$, sd = 3, $p = .00$).

The Rasch analysis can also provide "infit" and "outfit" statistical values for the facets of the research. For these two values, the quality control limit was set between 0.6 and 1.4 (Wright & Linacre, 1994). While the infit index gives information about the weighted squares average of the statistical fitness, the outfit index shows more sensitivity to outliers (Zhu, Ennis, & Chen, 1998). In other words, the infit index is expressed as information for unexpected answers in the decision-making process, while the outfit index expressed information for unexpected answers' outliers (Baştürk, 2010). The fact that both index values are around 1 (quality control values

1.4 - 0.6 range) means there was data suitability (Zhu et al., 1998). Table 2 proved that there is no value which exceeds the determined limit for both indices, given that both indices are between the determined values. Therefore, it can be specified that the infit and outfit values are included in the expected quality control values in the analysis of evaluating the HSEC.

The Rasch analysis results were compared with the SPSS results collected from the participants' views, and the general arithmetic means on the basis of grade variation are provided in Appendix 2. The data in the table indicates that the general arithmetic means of the 9th ($\bar{x}$ = 3.13±.61), 10th ($\bar{x}$ = 2.87±.52), 11th ($\bar{x}$ = 3.02±.63) and 12th ($\bar{x}$ = 2.82±.71) grades were found in the level of "sometimes". The general arithmetic means on the basis of grade variation show that the 12th grade curriculum had the lowest score while the 9th grade had the highest score in terms of suitability to the CIPP model.

Analysis of judges. The leniency or severity of judges of the HSEC according to the CIPP Model is presented in Table 3. When the rank of judges from the severest to the most lenient was considered, it was apparent that the judge coded J15 was the "most lenient" with 403 points and the judge coded J10 was the "severest" with 216 points. In addition, according to the numbers given to judges, the ranking from the most lenient to the severest is presented in Table 3 from top to bottom. It can be observed that there is a statistically significant difference between the severity/leniency of the judges (x^2 = 698.8, sd = 26, p = .00), as the separation index of the judges is 5.21 and the reliability coefficient is 0.96.

An examination of the infit and outfit statistical values of the facets shows that not all judges are among the expected quality control values (1.4 - 0.6 range). It appears that some of the judges are out of the expected values of the infit square average (J19, J4, J3, J12, J1) and some are out of the expected values of the outfit square average (J23, J24, J19, J4, J3, J12, J1). The fact that the infit and outfit square averages of the judges are in the expected values means there is consistency in the judges' scoring (Wright & Linacre, 1994). The above finding can be interpreted as the fact that some judges (J19, J4, J3, J12, J1, J23, J24) have not shown consistent scoring behaviors in the evaluation of HSEC according to the CIPP Model.

Table 3

A Comparison of Judges' Severity/Leniency

Obsvd Score	Obsvd Count	Obsvd Average	Fair Avrage	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Nu	Judges
403	100	4.0	4.06	1.26	.12	0.6	-3	0.7	-2	15	J15
384	100	3.8	3.86	.99	.12	1.6	3	1.6	3	23	J23
361	100	3.6	3.63	.69	.11	1.3	1	1.2	1	25	J25
343	100	3.4	3.45	.47	.11	0.6	-3	0.7	-3	2	J2
338	100	3.4	3.40	.41	.11	1.9	5	1.9	5	24	J24
335	100	3.3	3.36	.37	.11	0.9	0	1.0	0	22	J22
330	100	3.3	3.31	.32	.11	1.1	0	1.1	0	17	J17
324	100	3.2	3.25	.25	.11	1.3	2	1.3	2	11	J11
311	100	3.1	3.12	.10	.11	0.9	0	0.9	-1	7	J7
307	100	3.1	3.08	.05	.11	0.5	-5	0.5	-5	19	J19
306	100	3.1	3.07	.04	.11	0.6	-3	0.6	-3	8	J8
305	100	3.0	3.06	.03	.11	0.4	-6	0.4	-6	4	J4
297	100	3.0	2.97	-.06	.11	0.8	-1	0.9	-1	20	J20
289	100	2.9	2.89	-.15	.11	0.6	-3	0.6	-3	9	J9
284	98	2.9	2.89	-.15	.11	1.3	1	1.3	1	14	J14
288	100	2.9	2.88	-.16	.11	0.4	-6	0.4	-6	3	J3
287	100	2.9	2.87	-.17	.11	1.3	2	1.3	2	21	J21
271	100	2.7	2.71	-.35	.11	1.3	2	1.3	2	13	J13
269	100	2.7	2.68	-.38	.11	1.3	1	1.3	1	5	J5
263	100	2.6	2.62	-.44	.11	1.3	2	1.3	2	18	J18
251	100	2.5	2.50	-.58	.11	0.6	-3	0.6	-3	6	J6
235	100	2.3	2.33	-.77	.11	1.0	0	1.0	0	16	J16
234	100	2.3	2.32	-.79	.11	2.7	9	2.6	8	12	J12
233	100	2.3	2.31	-.80	.11	0.3	-8	0.3	-7	1	J1
221	100	2.2	2.19	-.95	.11	0.8	-1	0.8	-2	27	J27
219	100	2.2	2.17	-.97	.11	0.8	-1	0.8	-2	26	J26
216	100	2.2	2.14	-1.01	.11	0.8	-1	0.8	-1	10	J10
292.7	99.9	2.9	2.93	-.10	.11	1.0	-0.6	1.0	-0.6	Mean(Count:27)	
49.2	0.4	0.5	0.51	.58	.00	0.5	3.8	0.5	3.7	S.D.	

RMSE (Model) .11 Adj S.D. .57 Separation 5.21 Reliability .96
Fixed (all same) chi-square: 698.8 d.f.: 26 significance: .00
Random (normal) chi-square: 25.9 d.f.: 25 significance: .41

In the process of comparing the Rasch analysis with the SPSS program outputs, the study aimed to use the SPSS program to analyse whether there was a meaningful difference among teachers' gender variables in terms of grade considering the CIPP factors of the measurement tool. The data related to the participants' views on the evaluation of the 9th grade HSEC regarding the context, input, process and product factors in terms of gender is presented in Table 4. In the t-test analysis of HSEC, a significant difference ($p < .05$) among the views the English teachers in terms of gender variables for the related factors (excluding process factor: $F = 6.955$; $p = .011$) was noted. Female teachers ($\bar{x}_C = 3.63$, $\bar{x}_I = 3.33$, $\bar{x}_{Pc} = 3.46$, $\bar{x}_{Pd} = 3.12$) compared to male teachers ($\bar{x}_C = 3.21$, $\bar{x}_I = 2.71$, $\bar{x}_{Pc} = 3.19$, $\bar{x}_{Pd} = 2.54$) have shown a higher level of involvement in agreeing that the 9th grade curriculum has related factors.

However, when the analysis results of the 10th grade HSEC are examined, there is apparent that there is a significant difference among the participants' level of participation in terms of gender variation in the input and product factors of Stufflebeam's model ($p < .05$). It is apparent that the related difference is in favor of female teachers ($\bar{x}_I = 3.03$, $\bar{x}_{Pd} = 2.92$). On the other hand, while there was no significant difference in the context and process factors ($p > .05$), the participation level of male teachers ($\bar{x}_C = 3.00$) in the context factor and of female teachers (S = 3.21) in the process factor was higher.

Table 4
Teachers' Views towards HSEC in terms of Gender Variable

Curr.	Factors	Male-1 (n=30;%46.9)		Female-2 (n=34;%53.1)		*t* and *p* Values		Levene Test	
		$\bar{x}$	*ss*	$\bar{x}$	*ss*	*t*	*p*	*F*	*p*
9th Curr.	Context	3.21	.64	3.63	.58	-2.715	.009	.034	.854
	Input	2.71	.75	3.33	.95	-2.917	.005	.555	.459
	Process	3.19	.36	3.46	.53	-2.325	.023	6.955	.011
	Product	2.54	.65	3.12	.84	-3.112	.003	.413	.523
10th Curr.	Context	3.00	.62	2.90	.66	.643	.523	.248	.621
	Input	2.57	.69	3.03	.71	-2.623	.011	.530	.469
	Process	3.13	.38	3.21	.54	-.664	.509	1.373	.246
	Product	2.42	.68	2.92	.77	-2.754	.008	.001	.979
11th Curr.	Context	2.83	.59	3.57	.66	-4.636	.000	1.239	.270
	Input	2.55	.76	3.25	.67	-3.922	.000	.063	.803
	Process	3.09	.36	3.34	.58	-2.013	.048	5.564	.022
	Product	2.31	.65	3.27	.67	-5.805	.000	.014	.906
12th Curr.	Context	2.80	.88	3.14	.87	-1.536	.130	.195	.660
	Input	2.46	.75	3.25	.93	-3.696	.000	1.468	.230
	Process	2.98	.50	2.91	.69	.408	.685	.972	.328
	Product	2.35	.76	2.82	1.01	-2.069	.043	3.412	.070

$p < .05$.

Table 4 demonstrates that there was a significant difference among teachers' participation levels in terms of gender variable in the context, input and product factors of the 11th grade HSEC ($p < .05$). The related finding that the agreement levels of female teachers to the question of the 11th grade HSEC's compliance with the factors above ($\bar{x}_C = 3.57$, $\bar{x}_I = 3.25$, $\bar{x}_{Pd} = 3.27$) were higher than that of male teachers ($\bar{x}_C = 2.83$, $\bar{x}_I = 2.55$, $\bar{x}_{Pd} = 2.31$) can be excluded. When the level of teachers' agreement to the context, input, process and product factors of 12th grade HSEC was examined in terms of the gender variable, there was a significant difference among the agreement levels of teachers in the input and product factors ($p < .05$). Therefore, female teachers' participation levels ($\bar{x}_C = 3.25$, $\bar{x}_{Pd} = 2.82$) in the idea of the 12th grade HSEC's suitability to the input and product factors was higher compared to male teachers ($\bar{x}_I = 2.46$, $\bar{x}_{Pd} = 2.35$).

Analysis of scale items used in the evaluation of HSEC according to CIPP model. The appropriateness of the scale items used by the teachers for evaluating HSEC in the context of the CIPP model is presented in Table 5. The statistics related to the analysis of the difficulty of the items show that the separation index is 3.76 and the reliability coefficient is .93. The reliability coefficient (.93) indicates that the criteria used to determine the qualities of HSEC in the context of the CIPP model are reliable. It is also apparent that there is a significant difference among the difficulties of the items on the scale ($x^2 = 361.3$, sd = 24, $p = .00$). In this context, the most difficult item was the item coded II7, which is the input factor "Course books can draw students' attention." On the other hand, the easiest item was that coded CI6

"The project subjects are compatible with the course subjects", which is a context factor. Additionally, the standard error (RMSE) for the analysis of the scale items was .11. This value can be interpreted as quite low for determining the quality of the HSEC. The value of the adjusted standard deviation (Adj S.D. = .39) was below the critical value of 1.0 when considering this error rate.

There were no data that exceeded the limit of the determined values when the infit and outfit statistical values related to the research facets were considered. This demonstrates the consistency of the items used in the evaluation of the HSEC. In other words, it can be said that the infit and outfit mean squares of all of these items have acceptable values.

Table 5

Proficiency Statistics Used in the Evaluation of HSEC

Obsvd Score	Obsvd Count	Obsvd Average	Fair Avrage	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Nu	Items
253	108	2.3	2.31	.70	.11	1.2	1	1.2	1	7	II7
255	108	2.4	2.33	.68	.11	1.1	0	1.0	0	25	PdI25
265	107	2.5	2.45	.54	.11	0.8	-1	0.8	-1	23	PdI23
288	108	2.7	2.65	.31	.10	0.9	0	0.9	0	11	II11
289	108	2.7	2.66	.30	.10	1.1	0	1.1	0	3	CI3
289	108	2.7	2.66	.30	.10	1.1	0	1.0	0	17	PdI17
290	108	2.7	2.67	.28	.10	1.0	0	1.0	0	20	PdI20
291	108	2.7	2.68	.27	.10	0.8	-2	0.8	-2	24	PdI24
299	108	2.8	2.76	.19	.10	1.1	0	1.1	0	12	PcI12
302	108	2.8	2.79	.16	.10	0.7	-2	0.7	-2	21	PdI21
300	107	2.8	2.80	.15	.10	0.6	-3	0.6	-3	22	PdI22
305	108	2.8	2.82	.12	.10	1.1	0	1.1	0	2	CI2
307	108	2.8	2.84	.10	.10	0.6	-3	0.6	-3	19	PdI19
316	108	2.9	2.93	.01	.10	1.1	0	1.1	0	16	PcI16
317	108	2.9	2.94	-.01	.10	0.8	-1	0.8	-1	18	PdI18
319	108	3.0	2.96	-.03	.10	1.2	1	1.2	1	10	II10
331	108	3.1	3.08	-.16	.10	0.7	-2	0.7	-2	8	II8
338	108	3.1	3.15	-.23	.10	1.1	0	1.2	1	1	CI1
338	108	3.1	3.15	-.23	.10	1.3	2	1.3	1	5	CI5
345	108	3.2	3.21	-.31	.10	1.2	1	1.2	1	13	PcI13
352	108	3.3	3.28	-.38	.10	1.4	2	1.4	2	4	CI4
363	108	3.4	3.39	-.50	.11	1.0	0	1.0	0	9	II9
364	108	3.4	3.40	-.52	.11	1.1	0	1.1	0	14	PcI14
385	108	3.6	3.60	-.76	.11	1.0	0	1.0	0	15	PcI15
403	108	3.7	3.77	-.97	.11	1.2	1	1.1	0	6	CI6
Obsvd Score	Obsvd Count	Obsvd Average	Fair Avrage	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Nu	Items
316.2	107.9	2.9	2.93	.00	.11	1.0	-0.1	1.0	-0.1	Mean (Count: 25)	
37.4	0.3	0.3	0.36	.41	.00	0.2	1.7	0.2	1.7	S.D.	

RMSE (Model) .11 Adj S.D. .39 Separation 3.76 Reliability .93
Fixed (all same) chi-square: 361.3 d.f.: 24 significance: .00
Random (normal) chi-square: 23.9 d.f.: 23 significance: .41

Input: I Content: C Process: Pc Product: Pd

Derived from the SPSS program outputs performed to support the analysis of the Rasch model in terms of research factors and items, the means and standard deviations of teachers' views regarding the evaluation of HSEC according to the CIPP model are provided in Appendix 2. In the analysis, the sixth item in the context factor has the highest participation level with "mostly agree" among all of the grades ($\bar{x}_{9th\ grade}$ = 3.84±.84; $\bar{x}_{10th\ grade}$ = 3.76±.94; $\bar{x}_{11th\ grade}$ = 3.62±.98; $\bar{x}_{12th\ grade}$ = 3.64±.95). On the other hand, of the input factors the following had the lowest level with "rarely": the seventh item in the 9th grade ($\bar{x}$ = 2.41±1.18); the 25th item in the 10th ($\bar{x}$ =

2.31±1.08) and 11th grades ($\bar{x}$ = 2.37±1.01); the third item in the 12th grade ($\bar{x}$ = 2.34±1.18). When the arithmetic mean of the factors within the scope of analyses are examined, it is apparent that the product factor in the 9th ($\bar{x}$ = 2.85±.81), 10th ($\bar{x}$ = 2.69±.76), 11th ($\bar{x}$ = 2.82±.81) and 12th grades ($\bar{x}$ = 2.60±.93) was accepted at the lowest level with "rarely". In the context factor, the 9th grade ($\bar{x}$ = 3.44 ± .64) has the level of "mostly agree" while the 11th ($\bar{x}$ = 3.25±.72) and 12th ($\bar{x}$ = 2.97±.89) grades were "sometimes". Finally, in the process factor the 10th grade ($\bar{x}$ = 3.17±.47) had the highest level of participation with "sometimes" (Appendix 2).

Interaction analysis of judges' bias. The interaction analysis of the judges' views on HSEC regarding the CIPP Model is presented in Table 6. This data shows that the J11 has given 63 points to the 9th HSEC1, instead of 85 points, which means that this judge behaved in a severe manner. However, it is noted that the same judge has given 100 points instead of 76 to the 12th HSEC4 and here behaved leniently. Nonetheless, Z scores lying outside the range of +2 and -2 mean there is an interaction bias (Semerci, 2011), and in the present study Z scores are seen to range between -7.54 and 4.59.

Table 6
Interaction Analysis of HSEC with Judges

Obsvd Score	Exp. Score	Obsvd Count	Obs-Exp Average	Bias+ Model Measure	S.E.	Z-Score	Infit MnSq	Outfit MnSq	Sq	N	HSEC	measr	Nu	Judges	measr
63	84.8	25	-.87	.99	.22	4.59	0.9	0.9	41	1	9.HSEC 1	.17	11	J11	.25
30	50.1	25	-.80	1.89	.46	4.11	0.9	0.9	40	4	12.HSEC 4	-.21	10	J10	-1.01
67	86.2	25	-.77	.88	.21	4.11	0.4	0.4	65	1	9.HSEC. 1	.17	17	J17	.32
61	79.1	25	-.72	.82	.22	3.78	0.9	0.9	42	2	10.HSEC 2	-.09	11	J11	.25
46	62.8	23	-.73	.89	.25	3.62	1.4	1.5	56	4	12.HSEC. 4	-.21	14	J14	-.15
47	63.2	25	-.65	.84	.25	3.40	0.5	0.5	52	4	12.HSEC 4	-.21	13	J13	-.35
41	56.7	25	-.63	.93	.28	3.38	0.8	0.8	46	2	10.HSEC 2	-.09	12	J12	-.79
47	62.1	25	-.60	.79	.25	3.19	3.4	3.1	45	1	9.HSEC 1	.17	12	J12	-.79
59	73.1	25	-.57	.65	.22	2.95	0.6	0.6	28	4	12.HSEC 4	-.21	7	J7	.10
61	74.9	25	-.56	.63	.22	2.92	0.7	0.7	55	3	11.HSEC 3	.12	14	J14	-.15
58	70.0	25	-.48	.55	.22	2.51	1.3	1.3	19	3	11.HSEC 3	.12	5	J5	-.38
69	80.6	25	-.46	.52	.21	2.47	0.4	0.4	66	2	10.HSEC 2	-.09	17	J17	.32
72	82.6	25	-.43	.48	.21	2.28	2.3	2.3	94	2	10.HSEC 2	-.09	24	J24	.41
91	81.6	25	.38	-.45	.22	-2.01	0.9	0.8	25	1	9.HSEC 1	.17	7	J7	.10
75	65.3	25	.39	-.44	.21	-2.06	0.7	0.7	18	2	10.HSEC 2	-.09	5	J5	-.38
82	71.1	25	.44	-.49	.21	-2.30	0.9	0.9	17	1	9.HSEC 1	.17	5	J5	-.38
83	70.2	25	.51	-.58	.22	-2.68	0.6	0.6	54	2	10.HSEC 2	-.09	14	J14	-.15
103	89.4	25	.55	-.76	.25	-2.98	1.4	1.4	5	1	9.HSEC 1	.17	2	J2	.47
100	83.7	25	.65	-.84	.24	-3.46	0.5	0.5	43	3	11.HSEC 3	.12	11	J11	.25
94	76.1	25	.72	-.86	.23	-3.73	1.0	1.0	53	1	9.HSEC 1	.17	14	J14	-.15
79	57.4	25	.86	-.99	.21	-4.66	0.5	0.5	37	1	9.HSEC 1	.17	10	J10	-1.01
100	76.4	25	.94	-1.17	.24	-4.81	0.3	0.3	44	4	12.HSEC 4	-.21	11	J11	.25
97	70.5	25	1.06	-1.27	.24	-5.39	0.9	1.0	51	3	11.HSEC 3	.12	13	J13	-.35
105	78.0	25	1.08	-1.42	.26	-5.43	0.4	0.4	68	4	12.HSEC 4	-.21	17	J17	.32
99	61.0	25	1.52	-1.82	.24	-7.54	0.9	1.0	47	3	11.HSEC 3	.12	12	J12	-.79
Obsvd Score	Exp. Score	Obsvd Count	Obs-Exp Average	Bias+ Model Measure	S.E.	Z-Score	Infit MnSq	Outfit MnSq	Sq	N	HSEC	measr	Nu	Judges	measr
73.2	73.2	25.0	.00	.01	.22	.01	0.9	0.9	Mean (Count: 108)						
16.1	12.7	0.2	.39	.50	.03	2.02	0.9	0.5	S.D.						

Fixed (all = 0) chi-square: 442.4 d.f.: 108 significance: .00

When Table 6 is examined in detail in terms of bias, it can be said that the judges have exhibited extreme, severe or lenient behaviors in evaluating the HSEC. In the interaction analysis conducted by the judges for the evaluation of HSEC, the J10 has awarded 30 points (Z = 4.11) to the 12th HSEC4 whereas he should have given 50 points, and the same judge gave 79 points (Z = 4.66) to the 9th HSEC1 instead of 57 points, which means that the judge's behavior was biased. It is also apparent that the judge coded J12 gave 47 and 41 points, respectively, to the 9th and 10th HSEC,

whereas it was expected that these would be awarded 62 and 56 points ($Z_{9th\ grade}$ = - 3.19, $Z_{10th\ grade}$ = - 3.38). These findings can be interpreted as demonstrating that the relevant judges acted with severe or lenient behaviors while scoring the HSEC.

Findings Related to Qualitative Data

The themes obtained through the Maxqda program are presented in this section, which considers the questions on the interview form. The findings of the qualitative dimension of the research are collected under four themes: Content, Input, Process and Product. The participants' views on the HSEC are examined in detail and analyzed under each theme in the form of current situation, deficiencies and suggestions related to the HSEC, and presented in the form of models.

Findings related to the context factor of HSEC. The teachers' views about the current situation, deficiencies and suggestions on the basis of the context factor of HSEC are presented in Figure 2. This shows that certain views regarding the current state of the HSEC can be stated as follows: "Ambiguous state of attainment goals," "Unbalanced distribution of attainment goals among classes." The relevant codes were based on the views expressed by the participant coded $16P_M$ as follows, *"... when we look at the distribution of subjects in the 12th grade classes, where the lesson processes are not very effective, it can be seen that the topics are too dense and are not very suitable to the class hours*".

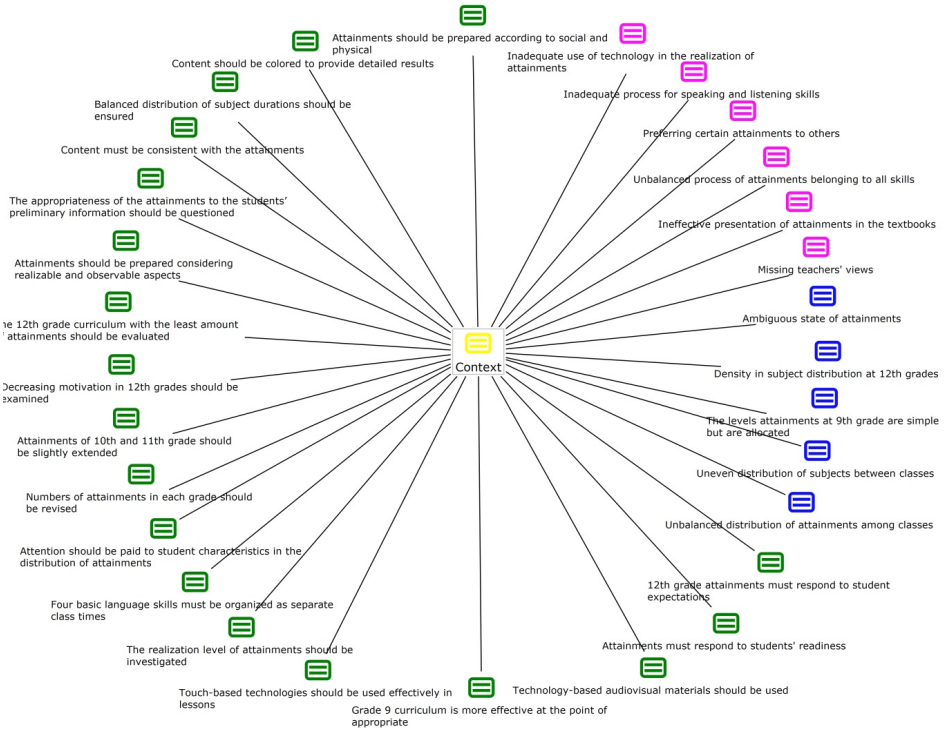


Figure 2. Participants' views on the context factor of HSEC.

In relation to the context factor, certain problems were also recorded in the evaluation of HSEC, such as "Unbalanced process of attainment goals belonging to all skills," "Inadequate use of technology in the realization of attainment goals." The feeling expressed by $4P_F$ was, *"…among technological materials, smart boards that are used most practically in classrooms present many useful facilities such as listening, reading, presentation of grammatical information and storytelling; unfortunately, I believe that not all teachers are using as much of the technology as possible in the realization of the curriculum attainment goal."* This has been used as a source to generate the related codes. Additionally, some of the other suggestions put forward were, "Technology-based audio-visual materials should be used," "The appropriateness of the attainment goal to the students' preliminary information should be questioned," "Four basic language skills must be organized as separate class times," and "Content should be colored to provide detailed results." The following opinions stated by $3P_F$ and $1P_M$ were also used to reference the codes, "*In classrooms the number of separate hours for the basic four language skills must be increased, because it is not possible to do activities for each skill in the current class hours. At least, listening skills can be applied during separate lesson hours and in this way it may be possible to evaluate listening skills. In this way, I think that students will*

be more willing to develop these skills by paying more attention to listening" and "*Touch-oriented technological tools such as smart boards should be examined to see how effectively they are being used in English classes*."

Findings related to the input factor of HSEC. The model created as a result of the content analysis of teachers' views regarding the input factor is presented in Figure 3. The codes generated are highlighted with different colors in the framework of the current situation, deficiencies and suggestions for the HSEC. In the process of interpretation of the codes, citations from the participants are also included in order to contribute to the reliability of the thematic dimension of the research. Following the examination of this data, there were twenty-two codes under the input theme. Some of the opinions which were related to the current situation of HSEC include, "Unbalanced theories and practices in the 12th grade," "As a follow-up to the curriculum, group decisions must be implemented effectively," and "Self-criticism: practitioners' disabilities in using existing possibilities." As a reference to these codes, the opinions of $14P_M$ are an example, "*Teachers' group decisions that are prepared in accordance with the inputs of the curriculum are not considered effectively... In teachers' meetings, listening and reading activities are mentioned as complementary materials of the curriculum. However, I see that the implementation and follow-up activities are not sufficient. It is also understood that plans considered in a similar way are not sufficiently functional.*"

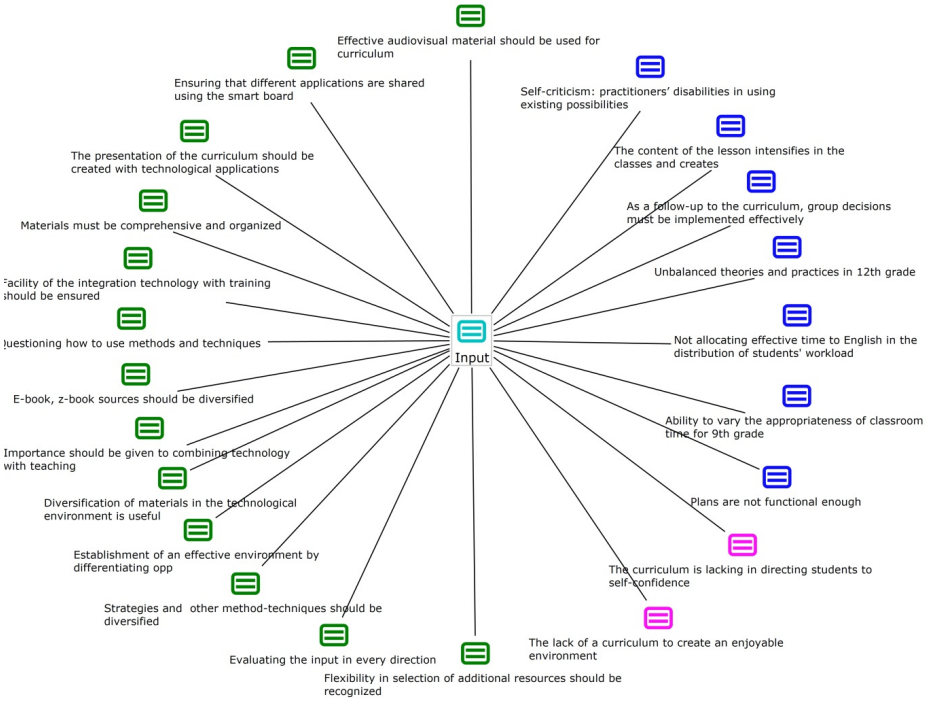


Figure 3. Participants' views on the input factor of HSEC.

In the context of the input factor, some of the deficiencies of the HSEC are indicated by views such as, "The lack of a curriculum to create an enjoyable environment," and "The curriculum does not enable students to develop self-confidence." These codes were created by taking into consideration the expression, "*Lack of activities that make students feel right in the centre of the lesson prevents the students from enjoying the lesson. Entertaining participation is not ensured. Students who cannot participate actively in the course may have shortcomings in self-confidence.*" Some of the views expressed within the scope of the suggestions can be represented as, "Strategies and other method-techniques should be diversified," "Facility of the integration technology with training should be ensured," and "E-book, z-book sources should be diversified." At this point, the expressions cited from the ideas of $4P_M$ and $3P_F$ were the basis of related codes and can be stated respectively as, "*The choice of additional materials should also be taken into account in the effective use of the curriculum and of the diversity of applications*" and "*The curriculum should provide the opportunity to combine technology with teaching applications. I also believe that we should give great importance to the integration of technology with the teaching environment.*"

Findings related to the process factor of HSEC. Figure 4 includes the codes regarding the views of teachers towards HSEC considering the process factor in Stufflebeam's CIPP model. Codes were generated based on the following statements about the current HSEC, "The application of the curriculum with classical understanding," "Ineffective presentation of content due to exams in 12th grades," and "The effectiveness of this curriculum should also be evaluated in university entrance exams." The related codes were based on the views expressed by $8P_F$ and $1P_M$ which were, "*Especially in the 12th grades, as the students prepare for university entrance exams, the content of the curriculum is not effective and is ignored. The most important reason for this is that the English course is not included in the exam courses. Students who have already entered for the exam with an intense amount of content are more interested in having question solutions and revision of subjects in the lessons that are not included in the exam"* ($8P_F$), and *"...For the awareness of seeing and questioning the efficiency of the curriculum, lessons which are not included in the university entrance exam such as English must be evaluated and the process examined.*" On the other hand, the views expressed with regard to the deficiencies of the HSEC were, "Inability to use student-centred activities such as group work," "Lack of English speaking," and "Equipping the sites with touch-sensitive materials." The related codes were based on the views expressed by $5P_F$ and $2P_F$ such as, "*The touch-based technology which makes it possible to use official site content like Eba on the smart board is inadequate*" ($5P_F$) and "*The lack of different applications, shortcomings in audio-visual materials, not using different methods and techniques often creates problems in learning language skills*" ($2P_F$).

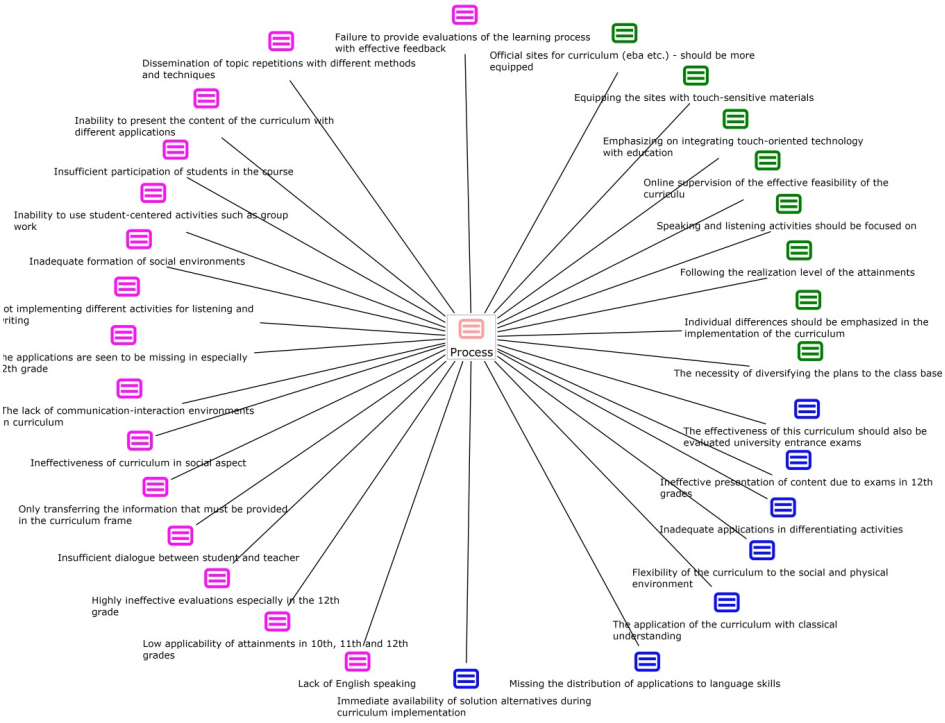


Figure 4. Participants' views on the process factor of HSEC.

The suggestions put forward regarding the deficiencies and problems related to the process factor include, "Individual differences should be emphasised in the implementation of the curriculum," "Equipping the sites with touch-sensitive materials," "Speaking and listening activities should be focused on," and "Official sites for curriculum (eba, etc.) should be better equipped." The related codes were generated based on the ideas of the participants coded as $4P_M$, $1P_M$ and $3P_F$ whose ideas are stated as follows: "*Generally, speaking and listening activities are not done in the class. Instead, grammatical exercises are presented and related examples solved on the board*" ($4P_M$); "*Of course, for an effective learning process, teachers need to use effective communication skills as much as possible. It should also be emphasized that individual differences should be considered at the same time*" ($1P_M$), and "*Equipping these types of sites with touch-based technology and using them instantly will make the process of the curriculum very effective and productive. It is currently used from time to time, but it is not enough to address the differences at the desired level. But I think it can be enriched and improved easily*."

Findings related to the product factor of HSEC. The final factor in the thematic aspect of the research is the acquisition of English teachers' views on the product factor of HSEC. In this respect, when the HSEC was evaluated according to the

product factor, some codes related to the current situation of the curriculum include, "Appropriateness of the content of the 9th grade course book to the social structure," "Smart boards enabling students and teachers to use self-evaluation," and "Seeing a moderate success of the curriculum." The abovementioned codes were created by referring to the ideas of participants $13P_M$ and $8P_F$. Their ideas included, "*When a general evaluation is made, the 9th grade course book is seen to be more consistent with the social structure and socio-cultural life and the content of it is consistent with the social structure in which the school is located"* ($13P_M$). "*It is absolutely necessary to take responsibility for oneself and to make evaluations in the framework of the school, province, region and country. Therefore, the curriculum should be prepared in a way to enable such evaluations through the network connected to the smart board and to enable self-evaluation of teachers and students*."

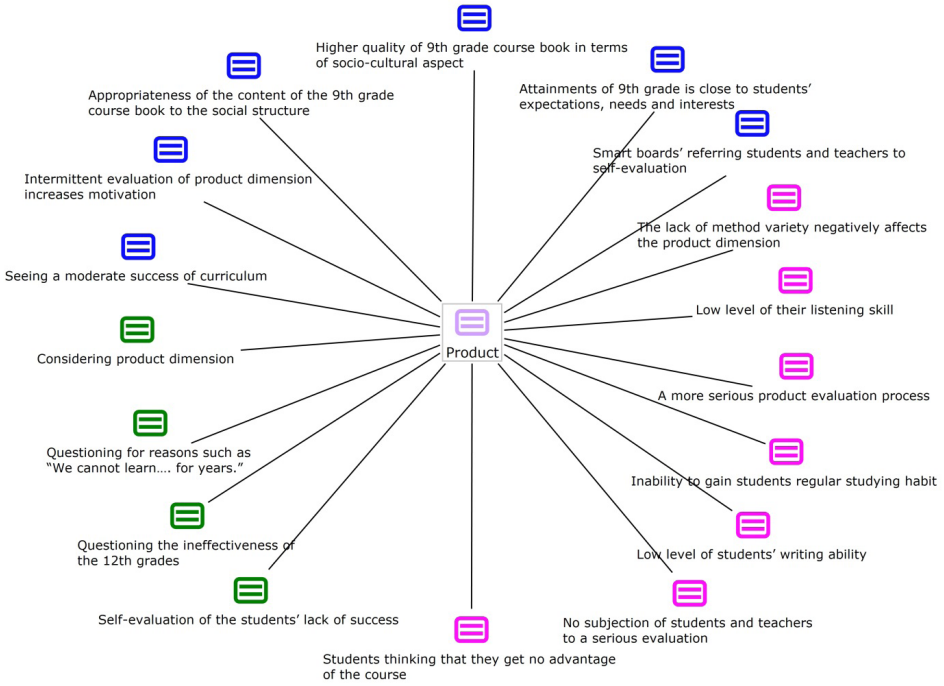


Figure 5. Participants' views on the product factor of HSEC.

In the context of the product factor, some of the deficient points of the HSEC were stated as, "Students thinking that they get no advantage from the course," "Low level of their listening skills," and "The lack of method variety negatively affects the product dimension." The related codes were based on the views expressed by $7P_M$ and $12P_F$ which can be highlighted as, *"...listening improvement is seen to be quite low. The reason for this deficiency is that students do not have the habit of acquiring these language skills in particular. Unlike this deficiency, it can be said that the*

things to be gained from grammar are a little better. It may be true that to determine great success cannot be achieved in the product dimension as the curriculum does not appeal to the students' interests and as there is no diversity of methods." In the research, suggestions about the deficiencies in the evaluation of HSEC according to the product factor are also reflected in the model. The codes such as, "Questioning the ineffectiveness of the 12th grades," "Questioning for reasons such as we cannot learn…. for years," and "Considering product dimension" are examples of the views highlighted by the participants. The related codes were based on the ideas of $9P_F$ whose ideas can be stated as, "*the questioning of students about the ineffective process in the class such as '…we have been trying to learn this foreign language for years, but we still cannot speak. We are not able to learn, we always learn similar topics' should be answered… and also product evaluation should be taken a bit more seriously, especially different studies and solutions suggested at this point."*

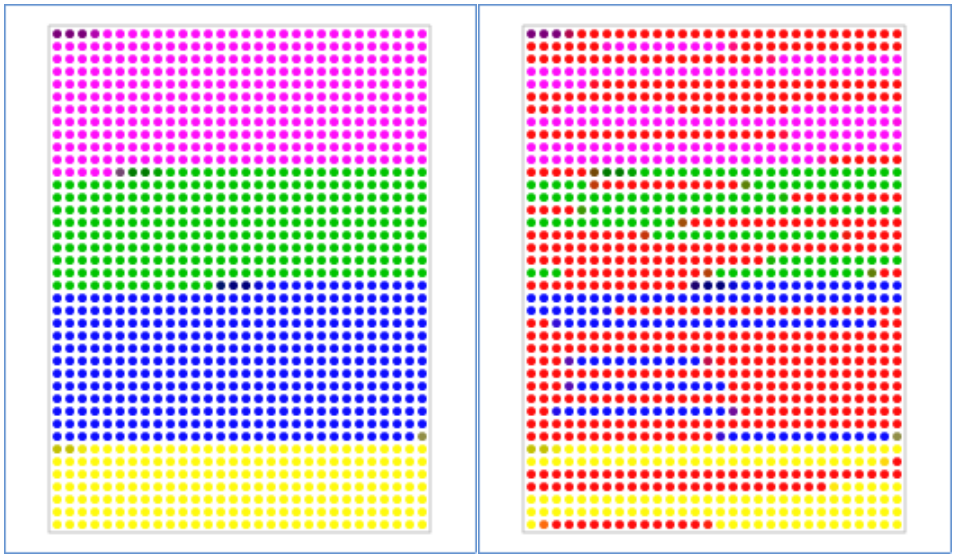


Figure 6. Document portrait of the codes and themes indicating English teachers' views regarding the evaluation of HSEC according to the CIPP Model. Pink: Content; green: input; blue: process; yellow: product; red: Deficiencies of themes regarding CIPP.

The document portrait in the Maxqda analysis enables the sequence and distribution of the codes in the study to be seen. Each color represents a different theme and each point represents a certain part of the coded parts. The size of the parts in the portrait (how much space they have occupied) is important and the colors are evaluated according to this magnitude (D'Andrea, Waters, & Rudd, 2011). When considering the code colors, there is a Porte consisting of 1.200 points in 30 lines and 40 points in each line.

In the present study, the views that participants indicated in the evaluation of the HSEC according to the CIPP model are shown under four different themes (context, input, process, product), so the codes including deficiencies for each theme are represented in red in the second portrait in Figure 6. Some of the codes regarding the deficiencies stated in the context dimension are, "Unbalanced process of subject matter belonging to all skills; preferring certain subject matter to others; missing teachers' views; ineffective presentation of subject matter in the textbooks". On the other hand, in the codes given under the input factor, the deficiencies represented indicated in the green color sequence appear to occupy less space than that of deficiencies in the context factor. Some of these deficiencies are as follows, "The lack of a curriculum to create an enjoyable environment and the curriculum does not direct students to self-confidence". The third theme in the portrait stated as process is shown in the blue color. Most of the codes under this blue color are recognized to be deficiencies, which is proven by the red colored points covering a large area of the second portrait in Figure 6. These red colored codes of the process factor may be stated as, "Lack of English speaking; inability to present the content of the curriculum with different applications; highly ineffective evaluations especially in the 12th grade; not implementing different activities for listening and writing; inadequate formation of social environments, and failure to provide evaluations of the learning process with effective feedback." The final theme in the portrait indicated in a yellow color is the product factor. The red color covers a certain part, which means that there are some deficiencies such as, "The lack of method variety negatively affects the product dimension; low level of their listening skills; no subjection of students and teachers to a serious evaluation; students thinking that they get no advantage from the course; inability of students to learn regular studying habit." Thus, it can be concluded that the most dominant part of the red color is occupied by the codes of the process factor, while the least dominant part of the red color is occupied by the input factor. The greatest deficiency of the HSEC according to the CIPP model is seen in the process factor, while it is least evident in the input factor.

Conclusion and Discussion

This research was conducted to obtain English teachers' perceptions with regard to the HSEC in the context of the CIPP model. In this respect, it is evident that the highest score for HSEC in the CIPP model, based on the results of the Rasch model, was in the 9th grade HSEC 1 coded curriculum, whereas the lowest score was in the 12th grade HSEC 4 coded curriculum. In the evaluation of all of the factors (context, input, process and product) in the grade levels in the SPSS analysis, the highest arithmetic mean was found in the 9th grade and the lowest was in the 12th grade (Appendix 2). According to the interpretation of the CIPP model, the 9th grade curriculum achieved the highest score and 12th grade curriculum had the lowest score. The fact that this result is the same as the Rasch analysis indicates that the

results gained from multiple analysis are consistent and complementary with each other. This increases the reliability of the research. On the other hand, the infit and outfit statistical values with regard to the program surface can be said to be between 0.6 and 1.4, which are expected quality control values. Thus, the result of the current research is different from the results of the study by Carvalho, Primi, and Meyer (2012), where the calculated infit and outfit values were not in the margins of the quality values. On the other hand, it is apparent that the infit and output values for the judges' score of performance in the process of the evaluation of students' basic skills in violin training in the study carried out by Baştürk (2012) were within the margins of the expected values, as were those of the current research. Therefore, the judges' consistent scoring performance in the former study is seen in the evaluation of all judges for HSEC, based on the CIPP model in the current research.

When the judges' scores were examined, it was found that the highest scores belonged to the judge coded J15; the scores of the judges coded J10 and J26 were at equal levels and their scoring performance was the lowest. In addition, it was found that the judges acted with bias as their scoring was different to that expected. When the infit and output statistical values in terms of the highest/lowest scoring were taken into account, it was demonstrated that not all of the judges performed consistently when scoring, as their scoring values were not in the margins of the quality control values (1.4-0.6). Furthermore, there were statistically significant differences among the judges' highest and lowest scoring performances ($x^2 = 698.8$, sd = 26, $p = .00$). It is thought that these differences could stem from the teachers' gender and working period variables. A similar result was seen in the study conducted by Sjaastad (2014), indicating that the judges acted differently in scoring in the process of the evaluation of mathematics training. These differences were suggested to be generated from the judges' proficiency in mathematics. In the current research, a t-test in the SPSS program was used to analyze whether there were significant differences in the teachers' evaluations for HSEC in grade and factor aspects with regard to the gender variable, based on the CIPP model. It was found that the differences generally took place in favor of women in regard to grades and some factor aspects according to the gender variable. This result can be interpreted as showing that the female teachers had more positive thoughts and acted sensitively on account of their structural characteristics. This case has similar results to the study conducted by Hamaideh (2010) indicating that female students' sensitivity and emotions are higher than male students. Similarly, some studies (Barrett, Robin, Pietromonaco, & Eyssell, 1998; Grossman & Wood, 1993) conducted at an international level show that women are more sensitive and emotional than men.

When the statistical values for the item difficulty analysis in the scale were examined, the scales to evaluate HSEC in the context of the CIPP model were reliable

according to the reliability coefficient values (.93). Further, the most difficult items in the scale were the items coded I7 and I25, whereas the easiest item was that coded I6. This result shows consistencies with the analysis results of SPSS conducted to support and complement the research and extend its scope in the multiple analysis. When the arithmetic means of the items for the curriculums for all the grades (9th, 10th, 11th and 12th) are considered, it is apparent that the item coded M7 has the lowest arithmetic mean and M6 has the highest arithmetic mean in all the grade levels.

The fact that similar results were obtained through different research methods increases the validity and reliability of the research. In this regard, Denzin (2010) and Patton (2002) asserted that the use of different methods and techniques can overcome the limited and biased results obtained from the use of one method and enhance the validity and the reliability of the study's results. Furthermore, Brewer and Hunter (1989, as cited in Türnüklü, 2001) emphasized that the use of a variety of data, such as the current study, will increase the study's validity. On the other hand, when the infit and outfit statistical values for the surface of the items in the scales were examined, it was evident that the data did not exceed the determined limit values. This result highlighted the consistent performance among the judges. Both index values were similarly found to be within the determined limits in the study conducted by Batdı and Elaldı (2016), which aimed to evaluate the materials used in the English teaching curriculum.

In the Rasch analysis results of this research, the reliability of the findings was considered. The data acquired from the Rasch model provided information about the measurement instrument from different angles (Engerlhard & Wind, 2013) and showed that the items in the scales used to evaluate the HSEC according to the CIPP model were at a good level with regard to expediency. In this regard, the reliability co-efficiency in respect of the determination of the quality of HSEC was found to be .93, the item quality .93 and the judges' highest/lowest performance scoring .96. Baştürk (2010) indicated that the closer these values are to +1.00, the greater their reliability in the interpretation of traditional reliability results. Therefore, the scale can be stated to be quite reliable for the evaluation of HSEC in respect of the CIPP model.

The teachers' perceptions were obtained through thematic analysis, in order to support, complement and combine the data acquired in the quantitative research, and perform the multiple analysis in the research. In relation to the context factor, which serves to collect the information necessary to plan and determine the attainment goals in the preparation of a curriculum (Özaydın, Günbatar, Önal, & Çakır, 2012), the shortcomings of the current curriculum, such as the attainment goals which are unknown or not distributed equally, or the distribution of subjects which are not arranged according to the grade levels were emphasized. It was seen that the use of technology-based materials at the maximum level, the creation of attainment

goals with respect to student, grade level and environmental factors, and the equal distribution of language skills were suggested to overcome these shortcomings. When the suggestions are taken into account, the attainment goals are understood to have been formulated in the context of the four basic skills for language. However, the prevailing opinion was that these skills are not distributed equally. At this point, it was mentioned that the curriculum prepared by the Ministry of National Education (MoNE) presented the four basic skills in each chapter in order to integrate and complement one another (MoNE, 2014). In this respect, the fact that the teachers were the implementers of the curriculum and should be a good mentor and model, fully fulfilling their responsibilities and duties, and acting in a selfless way, was suggested as a means to enable the process to be successfully completed by increasing the effectiveness of the materials. It was indicated in the thesis supervised by Tunç (2010) that in-service training activities are quite important to enable teachers to develop awareness of educational changes, developments and current language teaching methods; and contribute to teaching language through a communicative approach. On the other hand, it was stated in the thesis supervised by Wadden (1995) that the evaluation of the curriculum which was prepared for "Applied Sciences and Technology" with regard to the context dimension of Stufflebeam's model reviewed the curriculum from every respect. On the other hand, Hsieh (1999) evaluated the curriculum for the banking and insurance business technologies according to the CIPP model in his study and came to the conclusion that the curriculum was insufficient to meet the students' needs. This shows a similar result to that of the current research. In this case, it can be stated that the model can reveal the strengths and weaknesses of a curriculum by providing a detailed evaluation.

With regard to the input dimension, the current shortcomings mentioned included the fact that the plans were not quite functional, there was density of the content and an unequable distribution of the subjects. These can all bring about ineffectiveness in the teaching and learning process. In this respect, some suggestions to overcome the shortcomings were suggested, such as the use of different strategies considering school and student profiles or supplying a variety of materials supported with technology. This result shows similarities to the thesis research carried out by Burke-Sinclair (2002), which indicated that there was a lack of flexibility to use different strategies in the classroom in the evaluation of an adult curriculum through the CIPP model. Furthermore, the study conducted by Seçkin (2011) conveyed similar results to the current research in respect of the teachers' perceptions of the English course curriculum for the 4th grade. His research highlighted that it was not sufficiently appropriate in terms of content and the sequence of the subjects was not well-organized and consistent with each other. Based on these results, the process of creating input should be revised.

In the context of the process dimension, the suitability of the curriculum goals with the social environment, the insufficiencies in the distribution of language skills, and inadequacies in the applications and activities were revealed. These problems should be followed up to ascertain their extent. With regard to the product dimension, the study found that the curriculum was successfully implemented at a moderate level and the lack of methodological variety affects the product process negatively. It was recommended that greater importance should be given to this phase and the reason why there is a low success rate in language acquisition should be questioned.

As a result of the evaluation of the HSEC in all the grades (9th, 10th, 11th and 12th) through the CIPP model, it was found that the success level is at moderate and low levels and the distribution of the language skills are unequal. In this case, those who are responsible for designing and preparing the curriculum are advised to act with the student-centred principle in mind from a broader perspective than the process of determination and creation of goals. The research by Fritz (1996) stressed the importance of the CIPP model, indicating that it is an effective method which can be used to identify students' needs and interests. Therefore, it can be concluded that the detailed and comprehensive studies concerning teachers, students, parents and administrations' perceptions of the effectiveness of the curriculum should be implemented at certain intervals and the results of these studies should be reflected in the constitution of the new curriculum following serious analysis.

As a result of the evaluation of HSEC with regard to the CIPP model through the Rasch and SPSS analysis, the study found that the HSEC for the 9th grade had a high score and the 12th grade had a low score. These findings were also supported by the thematic analysis. The content analysis of the participants' perceptions revealed that the HSEC for the 9th grade is generally suitable and at a good level in relation to the attainment goals, content and student-centred aspects. However, the HSEC for the 12th grade was found to be ineffective in the attainment goals, evaluation and implementation aspects. In the evaluation of HSEC based on the teachers' perceptions, the most important reason given for the ineffectiveness of the curriculum for the 12th grade was that the students (students who study in the field of mathematics and science, literature and social sciences, mathematics and literature except the foreign language) in this grade focus on the university entrance exam and place less importance on the curriculum. The reasons why the curriculum was not fully implemented were that the English course is not part of the university entrance exam, and the course is not effectively provided to the students because they pay more attention to the other courses. Therefore, the teachers considered the ineffective implementation of the curriculum was a shortcoming.

On the other hand, the fact that students were not intensively involved in preparation for the university entrance exam and do not have exam anxiety may play a role in the

finding that the 9th grade HSEC had a higher score. This situation enables the teachers to develop their students' learning performance by implementing the curriculum in its entirety. The fact that the quality of the curriculum, according to the CIPP model dimensions, have similar results in the three analysis types shows that the research is consistent, reliable and valid. In consequence of the multiple analysis, the insufficient quality of the current HSEC and its stated inadequacies should remind the curriculum development experts of the need to take the necessary precautions. The purpose of the evaluation according to the CIPP model is not to reveal insufficiencies, but to develop the insufficient aspects of curriculums (Stufflebeam & Madaus, 2000). For example, arrangements can be made to resolve the unequal distribution of the language skills with regard to the attainment goals, and the insufficiencies of the curriculum such as ineffective group or cooperative works. In this respect, the research of Ersen Yanık (2007) in his thesis, which evaluated the curriculum for an English course, found that the four language skills did not provide the students with integrated understanding. McDonough and Shaw (1998) indicated that reading, writing, listening and speaking skills should be given in a balanced way, arguing that these skills ensure the perception of communication and interaction in real life at a good level. Different activities, materials and strategies can be employed to provide language skills in an integrated and balanced way in the learning process.

In contrast, the report for the Examination and Evaluation of the New Curriculums (2005) found that the factors such as student, teacher, material and context should be constantly monitored and unless these are implemented, it will be difficult for curriculums to change and develop. In addition, educators (teachers and administrators) should give prominence to the functional aspects of the curriculums and act in a more generous way by eradicating personal issues and undertaking their responsibilities sensitively. In this respect, Rolland's (2013) study, aimed at evaluating the curriculum for English courses, found that teachers' evaluations and feedback for the curriculum were quite important for curriculum development and their perceptions should be consulted. If feedback is taken into account it has a considerable effect on the success of the curriculum. In a study including the evaluation of a curriculum with the CIPP model, it was stated that the responsibilities of students, parents, society and the country can be realized through the detailed examination of every phase of the CIPP model. This will improve the quality of education (Patil & Kalekar, 2015).

This research which has conducted a multiple analysis, has emphasized the consistency and reliability of its results by comparison with similar results in different analyses. Therefore, it is thought that the comprehensible, reliable and valid results acquired through the evaluation of the curriculum using multiple analysis can enable other researchers to gain different perspectives. Their attention should be drawn to the use of multiple analysis as a different methodological process in future.

References

Akın, Ö., & Baştürk, R. (2012). Keman eğitiminde temel becerilerin Rasch Ölçme Modeli ile değerlendirilmesi [The evaluation of the basic skills in violin training by Many Facet Rasch Model]. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi, 31,* 175–187.

Aközbek, A. (2008). *The evaluation of 9th grade mathematics curriculum via the opinions of teachers and students by using context, input, process and product (CIPP) model (general high schools, vocational and technical high schools)* (Master's thesis, Yıldız Technical University, İstanbul, Turkey). Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Arıbaş, S., & Tok, H. (2004, July). *İlköğretimin birinci kademesinde yabancı dil öğretiminde karşılaşılan sorunların değerlendirilmesi* [Evaluating the problems encountered in foreign language instruction in the first level of elementary school]. Paper presented at the XIII. Ulusal Eğitim Bilimleri Kurultayı, İnönü Üniversitesi, Eğitim Fakültesi, Malatya, Turkey.

Ball, S. (2011). *Evaluating educational programs. ETS R&D scientific and policy contributions series.* A research report. Retrieved September 5, 2016 from https://www.ets.org/Media/Research/pdf/RR-11-15.pdf

Bamberger, M. (2012). Introduction to mixed methods in impact evaluation. *Impact Evaluation Notes*, *3*, 1–38.

Barrett, L. F., Robin, L., Pietromonaco, P. R., & Eyssell, K. M. (1998). Are women the "more emotional" sex? Evidence from emotional experiences in social context. *Cognition and Emotion, 12*(4), 555–578.

Baştürk, R. (2010). Bilimsel araştırma ödevlerinin çok yüzeyli Rasch Ölçme Modeli ile değerlendirilmesi [Evaluating the scientific research assignment through many-facet Rasch Measurement]. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 1*(1), 51–57.

Batdı, V., & Elaldı, Ş. (2016). Analysis of High School English curriculum materials through Rasch Measurement Model and Maxqda. *Educational Sciences: Theory & Practice, 16*, 1325–1347. http://dx.doi.org/10.12738/estp.2016.4.0290

Bauer, C. F. (2008). Attitude toward chemistry: A semantic differential instrument for assessing curriculum impacts. *Journal of Chemical Education*, *85*(10), 1440–1445.

Boone, W., & Scantlebury, K. (2006). The role of Rash analysis when conducting science education research utilizing multiple-choice tests. *Science Education, 90*, 253–269.

Braun, V., & Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. Retrieved from http://eprints.uwe.ac.uk/11735

Brewer, E. W. (2010). Evaluation models for evaluating programs. In C. Victor & X. Wang (Eds.), *Assessing and evaluating adult learning in career and technical education* (pp. 106–126). Hangzhou, China: Zhejiang University Press.

Burke-Sinclair, M. (2012). *Utilizing Stufflebeam's CIPP model to evaluate an adult degree completion program.* (Doctoral dissertation). Available from ProQuest Dissertation and Theses database. (UMI No.3541477)

Büyüköztürk, Ş., Kılıç-Çakmak, E., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2009). *Bilimsel araştırma yöntemleri* [Scientific research methods] (3th ed.). Ankara, Turkey: Pegem Yayıncılık.

Carvalho, L. F., Primi, R., & Meyer, G. J. (2012). Application of the Rasch model in measuring personality disorders. *Trends Psychiatry Psychother, 34*(2), 101–109.

Clewell, B. C., Cohen, C. C., Campbell, P. B., Deterding, N., Manes, S., Tsui, L., … Carson, R. (2004). *Review of evaluation studies of mathematics and science curricula and professional development models*. Retrieved September 02, 2016 from http://www.urban.org/UploadedPDF/411149.pdf

Cook, D. A. (2010). Twelve tips for evaluating educational programs. *Medical Teacher, 32,* 296–301.

Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative.* New Jersey, NJ: Upper Saddle River.

Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed methods research designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 209–240). Thousand Oaks, CA: Sage.

D'Andrea, L. M., Waters, C., & Rudd, R. (2011). Using computer assisted qualitative software (CAQDAS) to evaluate a novel teaching method for introductory statistics. *International Journal of Technology in Teaching and Learning, 7*(1), 48–60.

Demirel, Ö. (2012). *Eğitimde program geliştirme* [Curriculum development in education] (18th ed.) Ankara, Turkey: Pegem Akademi.

Denzin, K. N. (2010). Moments, mixed methods, and paradigm dialogs. *Qualitative Inquiry, 16*(6), 419–427. http://dx.doi.org/ 10.1177/1077800410364608

Diamond, R. M. (1998). *Designing and assessing courses and curricula: A practical guide*. The Jossey-Bass Higher and Adult Education Series. San Francisco, CA: Jossey-Bass, Inc.

Duru, A., & Korkmaz, H. (2010). Öğretmenlerin yeni matematik programı hakkındaki görüşleri ve program değişim sürecinde karşılaşılan zorluklar [Teachers' views about a new mathematics curriculum and difficulties encountering curriculum change]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, *38*(38), 67–81.

Engelhard, G., & Wind, S. A. (2013). *Rating quality studies using Rasch measurement theory* (College Board Research Reports). Retrieved August 30, 2016 from https://research.collegeboard.org/sites/default/files/publications/2013/8/researchreport-2013-3-rating-quality-studies-using-rasch-measurement-theory.pdf

Ersen Yanık, A. (2007). *A study of English language curriculum implementation in 6th, 7th and 8th grades of public primary schools through teachers' and students' perceptions* (Doctoral dissertation, The Graduate School of Social Sciences, Middle East Technical University, Ankara, Turkey).

Examination and Evaluation of the New Curriculums. (2005). *This paper was 1 January 2017.* Retrieved from http://ilkogretim-online.org.tr/vol5say1/yenimufredat_raporu[1].pdf

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines*. New York, NY: Pearson Education.

Fritz, S. (1996). *Assessing undergraduate student needs utilizing the CIPP model of evaluation* (Doctoral dissertation). Retrieved from ProQuest Dissertation and Theses database. (UMI No. 9710261).

Gandomkar, R., Jalili, M., & Mirzazadeh, A. (2015). Evaluating assessment programmes using programme evaluation models, *Medical Teacher, 37*(8), 792–793. http://dx.doi.org/10.3109/0142159X.2015.1042436

Gökdemir, C. V. (2005). Üniversitelerimizde verilen yabancı dil öğretimindeki başarı durumumuz [Foreign language success at universities]. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 6*(2), 251–264.

Grossman, M., & Wood, W. (1993). Sex differences in intensity of emotional experience: A social role interpretation. *Journal of Personality and Social Psychology*, *65*(5), 1010–1022.

Hamaideh, S. H. (2010). Gender differences in stressors and reactions to stressors among Jordanian university students. *International Journal of Social Psychiatry, 58*(1), 26–33.

Hsieh, W. K. (1999). *Assessing the effectiveness of two-year banking and insurance technology programs of junior colleges in Taiwan, R.O.C.: An application of the context input process product (CIPP) evaluation model* (Doctoral dissertation). Available from ProQuest Dissertation and Theses database. (UMI No. 9936933)

Işık, A. (2008). Yabancı dil eğitimimizdeki yanlışlar nereden kaynaklanıyor? [Reasons of wrongs in foreign language teaching]. *Journal of Language and Linguistic Studies, 4*(2), 15–26.

Kali, Y., & Linn, M. (2010). Curriculum design – as subject matter: Science. In B. McGraw, E. Baker & P. Peterson Eds.), *International encyclopedia of education* (3rd ed., pp. 468–474). Philadelphia: Elsevier Science Pub.

Karasar, N. (2012). *Bilimsel araştırma yöntemi* [Scientific research method] (23rd ed.). Ankara, Turkey: Nobel Publication.

Karataş, H., & Fer, S. (2009). Evaluation of English Curriculum at Yıldız Technical University using CIPP model. *Education and Science, 34*(153), 47–60.

Karataş, H., & Fer, S. (2011). CIPP evaluation model scale: Development, reliability and validity. *Procedia Social and Behavioral Sciences, 15*, 592–599. http://dx.doi.org/10.1016/j.sbspro.2011.03.146

Klenowski, V. (2010). *Curriculum evaluation: Approaches and methodologies*. International Encyclopedia of Education. Elsevier Ltd. Retrieved September 06, 2016 from http://eprints.qut.edu.au/26163/1/Klenowski_curriculum_evaluation.pdf

Linacre, J. M. (1993, April). *Generalizability theory and many-facet Rasch measurement*. Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta.

Marsh, C. J., & Willis, G. (2007). *Curriculum: Alternative approaches, ongoing issues.* New Jersey, NY: Pearson Prentice Hall.

Marvasti, A. B. (2004). *Qualitative research in sociology.* London, UK: Sage.

McDonough, J., & Shaw, C. (1998). *Materials and methods in ELT.* Oxford, UK: Blackwell Publishers Inc.

Ministry of National Education. (2014). *High school English curriculum (9th, 10th, 11th, and 12th grades)*. Ankara, Turkey: Author.

Mirzazadeh, A., Gandomkar, R., Hejri, S. M., Hassanzadeh, G., Koochak, H. E, Golestani, A., … Razavi, S. H. (2016). Undergraduate medical education programme renewal: A longitudinal context, input, process and product evaluation study. *Perspectives on Medical Education, 5*(15), 15–23. http://dx.doi.org/10.1007/s40037-015-0243-3

Olsen, C., & George, D. M. M. (2004). *Cross-sectional study design and data analysis. College Entrance Examination Board.* Retrieved January 21, 2017 from http://www.collegeboard.com/prod_downloads/yes/4297_MODULE_05.pdf

Orbeyi, S., & Güven, B. (2008). Yeni İlköğretim Matematik Dersi Öğretim Programı'nın değerlendirme öğesine ilişkin öğretmen görüşleri [Teachers' views on evaluation dimension of Primary School new Mathematic curriculum]. *Eğitimde Kuram ve Uygulama*, *4*(1), 133–147.

Özaydın, B., Günbatar, M. S., Önal, N., & Çakır, H. (2012). Öğretmen yetiştirme programlarının sistemik ve sistematik değerlendirilmesi [Systemic and systematic evaluation of teacher training programs]. *Eğitim Teknolojisi Kuram ve Uygulama, 2*(1), 54–71.

Özdamar, K. (2010). *Paket programlar ile istatiksel veri analizi 2* [Statistical data analysis through package programs 2] (7th ed.). Eskişehir, Turkey: Kaan Publication.

Özdemir, M. (2009). Eğitimde program değerlendirme ve Türkiye'de eğitim programlarını değerlendirme çalışmalarının incelenmesi [Curriculum evaluation in education and examination of the curriculum evaluation studies in Turkey]. *Yüzüncü Yıl Üniversitesi, Eğitim Fakültesi Dergisi*, *6*(2), 126–149.

Parmaksız, R., & İncirci, A. (2016). 10. sınıf ortaöğretim İngilizce öğretim programının Stufflebeam (CIPP) bağlam, girdi, süreç ve ürün modeline göre değerlendirilmesi [Evaluating 10th grade English language curriculum using Stufflebeam's (CIPP) context, input, process, product model]. *International Journal of Language Academy, 4*(2), 153–173.

Patil, Y., & Kalekar, S. (2015). CIPP model for school evaluation. *Scholarly, Research Journal for Humanity Science & English Language, 2*(10), 2615–2619.

Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). London, UK: Sage. Retrieved December 25, 2016 from https://books.google.com.tr/books?id=FjBw2oi8El4C&printsec=frontcover&hl=tr#v=onepage&q&f=false

Rolland, K. A. (2013). *Evaluation of a school district's English as a second language program* (Doctoral dissertation). Available from ProQuest Dissertation and Theses database. (UMI No. 3551351)

Rolstad, K., Mahoney, K., & Glass, G. V. (2005). The big picture: A meta-analysis of program effectiveness research on English language learners. *Educational Policy, 19*(4), 572–594.

Sağlam, M., & Yüksel, İ. (2007). Program değerlendirmede meta-analiz ve meta-değerlendirme [Meta-analysis and meta-evaluation methods in program evaluation]. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi, 18,* 175–188.

Şahin, İ. (2007). Yeni ilköğretim 1. kademe Türkçe programının değerlendirilmesi [Assessment of New Turkish Curriculum for Grade 1 to 5]. *İlköğretim Online*, *6*(2), 284–304.

Seçkin, H. (2011). İlköğretim 4. sınıf İngilizce öğretim programına ilişkin öğretmen görüşleri [Teachers' views on primary school English language teaching curriculum for the 4th grade]. *Uluslararası İnsan Bilimleri Dergisi, 8*(2), 550–577.

Semerci, Ç. (2011). Mikro öğretim uygulamalarının çok yüzeyli Rasch ölçme modeli ile analizi [Analyzing microteaching applications with Many-facet Rasch Measurement Model]. *Eğitim ve Bilim*, *36*(161), 14–25.

Setyoko, A., Tunas, B., & Sunaryo, W. (2016). Evaluation of school operational assistance by using CIPP model in Indonesia private Islamic elementary school. *International Journal of Managerial Studies and Research, 4*(3), 44–49.

Sjaastad, J. (2014). Enhancing measurement in science and mathematics education research through Rasch analysis: Rationale and properties. *Nordina, 10*(2), 212–230.

Smith, B., & Benjamaporn, P. (2013). Evaluation of public health communication performance by Stufflebeam's CIPP model: A case study of Thailand's department of disease control. *Journal of Business and Behavioral Sciences, 25*(1), 36–53.

Steinert, Y., Cruess S., Cruess R., & Snell, L. (2005). Faculty development for teaching and evaluating professionalism: From programme design to curriculum change. *Medical Education, 39,* 127–36.

Stufflebeam, D. L. (1994). Empowerment evaluation, objectivist evaluation, and evaluation standards: Where the future of evaluation should not go and where it needs to go. *American Journal of Education, 15,* 321–338. http://dx.doi.org/10.1177/109821409401500313

Stufflebeam, D. L., & Madaus, G. F. (2000). Program evaluation: A historical overview. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (pp. 3–18). Springer. Retrieved from https://link.springer.com/book/10.1007/0-306-47559-6

Tseng, K. H., Diez, C. R., Lou, S. J., Tsai, H. L., & Tsai, T. S. (2010). Using the context, input, process and product model to assess an engineering curriculum. *World Transactions on Engineering and Technology Education, 8*(3), 256–261.

Tunç, F. (2010). *Evaluation of an English language teaching program at a public university using CIPP model* (Master's thesis, The Graduate School of Social Sciences, Middle East Technical University, Ankara, Turkey).

Türnüklü, A. (2001). Eğitimbilim alanında aynı araştırma sorusunu yanıtlamak için farklı araştırma tekniklerinin birlikte kullanılması [Using different research techniques for the same question in educational science]. *Eğitim ve Bilim, 26*(120), 8–13.

Varış, F. (1998). Temel kavramlar ve program geliştirmeye sistematik yaklaşım [Basic terms and a systhematic approach to curriculum development]. In A. Hakan (Ed.), *Eğitim bilimlerinde yenilikler* (pp. 3–19). Eskişehir, Turkey: Anadolu Üniversitesi Açık Öğretim Fakültesi Yayınları.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine, 37*(5), 360–363.

Wadden, M. M. (1995). *Development and implementation of a program monitoring model for Cabot college of applied ats, technology and continuing education using Stufflebeam's context evaluation* (Master's thesis, Memorial University, Newfoundland). Retrieved from http://research.library.mun.ca/4329/3/Wadden_MaryM.pdf

Wolf, P., Hill, A., & Evers, F. (2006). *Handbook for curriculum assessment*. Guelph, Ontario: University of Guelph Publications. Retrieved September 06, 2016 from http://www.tss.uoguelph.ca/resources/pdfs/HbonCurriculumAssmt.pdf

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. Transactions of the Rasch measurement SIG American Educational Research Association. *Rasch Measurement Transaction, 8*(3), 370–382.

Yin, R. K. (1994). *Case study research: Design and methods* (2nd ed.). Newbury Park, CA: Sage Publications.

Zehir Topkaya, E., & Küçük, Ö. (2010). An evaluation of 4th and 5th grade English language teaching program. *Elementary Education Online, 9*(1), 52–65.

Zhu, W., Ennis, C. D., & Chen, A. (1998). Many-faceted Rasch modeling expert judgment in test development. *Measurement in Physical Education and Exercise Science, 2*(1), 21–39.

Zikmund, W. G. (2003). *Research methods. Health economics research method, 2.* Retrieved August 28, 2016 from http://pioneer.netserv.chula.ac.th/~ppongsa/2900600/LMRM02.pdf

Appendix-1

Cohen Kappa Values of Themes in the Study

Context					Input					Process					Product				
		K2					K2					K2					K2		
		+	−	Σ			+	−	Σ			+	−	Σ			+	−	Σ
K1	+	29	2	31	K1	+	22	1	23	K1	+	30	3	33	K1	+	17	1	18
	-	3	22	25		-	2	17	19		-	2	24	26		-	0	8	8
	Σ	32	24	56		Σ	24	18	42		Σ	32	27	59		Σ	17	9	26
Kappa: .819 p:.000					Kappa: .855 p:.000					Kappa: .829 p:.000					Kappa:.913 p:.000				

Appendix-2

Mean and Standard Deviation Values of Items and Factors Related To Teachers' Assessment of HSEC (9th, 10th, 11th, 112th) According to CIPP Model

Factors	I.N	9th grade		10th grade		11th grade		12th grade	
		$\bar{x}$	ss	$\bar{x}$	ss	$\bar{x}$	ss	$\bar{x}$	ss
Context Asses.	1	3.50	.87	2.83	1.02	3.37	.98	3.03	1.19
	2	3.05	.93	2.41	.95	2.89	.89	2.78	1.21
	3	3.00	.99	2.39	.98	3.03	1.01	2.34	1.18
	4	3.67	1.11	3.10	1.03	3.33	1.10	3.20	1.30
	5	3.56	1.00	3.22	1.17	3.11	1.08	2.87	1.20
	6	3.84	.84	3.76	.94	3.62	.98	3.64	.95
Input Asses.	7	2.41	1.18	2.35	1.01	2.41	1.20	2.55	1.27
	8	3.12	1.08	2.97	1.00	3.23	1.08	3.05	1.03
	9	3.62	1.07	3.26	1.07	3.22	1.10	3.19	.81
	10	3.22	1.30	2.76	1.09	3.03	.99	2.87	1.36
	11	2.83	1.20	2.73	1.21	2.70	1.12	2.73	1.20
Process Asses.	12	2.75	.80	2.65	1.09	2.92	1.30	2.70	1.24
	13	3.51	.92	3.26	.84	3.16	1.04	2.91	.94
	14	3.58	.81	3.53	.85	3.37	.68	3.08	1.06
	15	3.69	.83	3.44	.77	3.56	.71	3.44	.75
	16	3.14	1.10	2.98	.98	3.11	.99	2.61	1.16
Product Asses.	17	2.64	1.22	2.75	1.04	3.01	1.33	2.59	1.22
	18	2.97	.83	3.16	1.06	2.94	.94	2.72	.93
	19	3.14	.89	2.83	.77	2.98	.79	2.69	1.10
	20	3.31	1.18	2.59	1.05	2.83	.98	2.50	1.19
	21	2.84	1.09	2.70	.97	3.14	1.11	2.69	1.07
	22	2.91	1.00	2.78	.82	2.69	.89	2.86	1.06
	23	2.58	.89	2.37	.88	2.62	.90	2.51	1.31
	24	2.64	1.16	2.70	.92	2.78	.93	2.56	1.08
	25	2.62	1.29	2.31	1.08	2.37	1.01	2.28	1.01
Total Arithmetic Means Related to Sub-Dimensions									
Factors	Context	3.44	.64	2.95	.64	3.25	.72	2.97	.89
	Input	3.04	.91	2.82	.73	2.92	.79	2.88	.93
	Process	3.33	.47	3.17	.47	3.22	.50	2.95	.61
	Product	2.85	.81	2.69	.76	2.82	.81	2.60	.93
$\bar{x}_{GAM}$		3.13	.61	2.87	.52	3.02	.63	2.82	.71

($\tilde{x}_{GAM}$ = : General Arithmetic Means of all items in the scale) (N:64)

Appendix-3

Scale Items Used in the Study

I.N	Items	I.N	Items
1	The duration and the topics in the curriculum are compatible.	14	The curriculum includes individual activities.
2	Theoretical information (book information) in the curriculum is sufficient.	15	Activities applied in the curriculum are boring.
3	Application of theoretical knowledge in the curriculum is sufficient.	16	The curriculum includes collaborative (group) activities.
4	The total duration of the curriculum is sufficient.	17	The curriculum responds to the individual needs of the students.
5	Time devoted to activities in the curriculum is sufficient.	18	The curriculum is the basis for students' foreign language needs in the future.
6	The project topics are compatible with course subjects.	19	The curriculum is suitable for students' development of foreign language skills.
7	The course book can attract students' attention.	20	The curriculum improves students' interpretation ability in foreign languages.
8	The topics in the course book are consistent with the goals of the curriculum.	21	The curriculum improves students' ability to think in foreign languages.
9	The topics included in the course book are understandable.	22	The curriculum improves the ability to relate foreign language and mother tongue.
10	The material used in the curriculum facilitates the learning of the students.	23	The curriculum improves students' ability to use foreign languages in other courses.
11	The materials used in the curriculum are of interest to the students.	24	The attainment goals (topics) of the curriculum meet the foreign language needs of the students.
12	The learning activities applied in the curriculum are teacher-centred.	25	The curriculum makes students enjoy learning foreign languages.
13	The learning activities applied in the curriculum are student-centred.		