Tania Blažević
Sveučilište u Splitu, Pravni fakultet
HR–21000 Split, Ulica Domovinskog rata 8
tblazevic@pravst.hr

Mara Ruža Blažević
Sveučilište u Splitu, Fakultet elektrotehnike, strojarstva i brodogradnje
HR–21000 Split, Ulica Ruđera Boškovića 32
mrblazev@fesb.hr

STUDENTS' PERCEPTION OF SECONDARY SCHOOL TEACHERS' ASSESSMENT WITH SPECIFIC REFERENCE TO THE SUBJECT OF ENGLISH

Abstract

Using research investigating the effect of teacher assessment practices in secondary schooling in Croatia on student perceptions of fairness, this paper[1] points out participating students' perception of fairness of assessment, focusing on teacher assessment in three secondary school subjects, but particularly in the subject of English. The research overall included enquiry into the fairness of teacher assessment practices in three secondary school subjects. The implemented research was both of a quantitative and qualitative nature. The research design was intended to be both empirical, measuring students' perception, and quasi-ethnographic, seeking to interpret these perceptions through understanding the context. Student questionnaires, involving a total of 330 students, teacher interviews, involving a total of 12 teachers, and analysis of teachers' assessment documents were used to collect data that were then analysed within case studies. Each case study, totalling six in number, represented a teacher with two classes of students studying one of the chosen three subjects. Classes of secondary school students from two high schools studying the subjects of Biology, Croatian or English as a foreign language formed the sample groups. These students were asked to complete a questionnaire based on Likert response type items founded on the Student Perception of Assessment Questionnaire (SPAQ) and open-ended questions based on a questionnaire developed at the Centre for Schooling and Learning Technologies (CSaLT). The SPAQ items were measured by five scales. The teachers of these students were interviewed and also provided documents that illustrate how they assessed students in these three subjects. The investigation uncovered variations in responses to teachers and subjects and differences in student response between classes of the same teacher which are possibly indicative, among other elements, of the role of teacher personality factors. This investigation revealed that from the three secondary school subjects researched, students' perception of the secondary school subject of English achieved the most positive rating among the sample of students involved in the research.

Key words: student perceptions, fairness, consistency, assessment, ESL

---

1   This paper makes reference to data from one of the author's doctoral thesis (see Blazevic, 2019).

1. Introduction

Perception has long been at the root of adages and people's attitude towards various elements in life. Related familiar sayings as 'beauty is in the eye of the beholder' and 'clothes do not make the man' and 'one man's rubbish is another man's treasure', amongst others, are often resounded. Perception by the Cambridge Dictionary (2003) is defined as "a belief or opinion, often held by many people and based on how things seem." Perceptual psychology goes on to purport that perception can have fundamental impact on behaviour which can in turn determine success and survival, shaping culture and society (Bruner 1960). An overview of existing research on the concept of perception and its relationship to assessment as well as its connection to knowledge, beliefs and attitudes, will be discussed in the subsequent section.

2. Perception and Assessment

One of the possible results of teacher assessment practices is student perception of teacher's assessment of student work. Here, understanding of student perceptions of teacher assessment bears attention. There appears to be numerous variations in perception among students of teacher assessment. The results from investigations carried out by Alkharusi (2015) indicate that students share a common experience regarding assessment resulting in a common student perception within a certain class about their teacher's assessment practices. Alkharusi goes on to state that individual student perceptions of teacher assessment should be considered, as well as the average perceptions of all the students in that particular class environment. Alkharusi then proceeds to classify that particular class environment as the objective environment despite the subjectivity of each student's perception. He states that his particular class or classroom environment is the environment which is believed to be real. For Alkaharusi, the perceived classroom assessment environment was recorded at a collective classroom level rather than at an individual level. Then, from this environment, there were differentiated class results arising from assessment practices of the teacher (Alkharusi 2015). That is, differential assessment practices were implemented resulting in differential data. For example, there were even different class perceptions based on the gender weighting of the class. In conclusion, Alkharusi believes that students' perceptions of the assessment environment of the classroom requires more precise measurement.

Nevertheless, where an individual student's perception of teacher assessment practices varies from that of the class as a whole, this individual student's perception could be downscaled or less relevant. This reflects adherence to the idea of *majority rules* where the view of most students of the teacher's assessment practices carries the most weight as being dominant and is in fact the dominantly most valid one. The individual student experience becomes that of the group. This varies however among classes. Needless to say, teacher assessment practices impact on this group experience (Alkharusi 2008). It is better to interpret the student experience as a collective rather than as a set of individual response. Alkharusi (2008) mentions that students' perception of the classroom environment will be more positive if there is better student-teacher communication in regards to classroom assessment.

This study revealed that the perceptions of student research participants related to assessment tasks all culminate in positively influencing student self-efficiency and motivation levels. This resulted in the conclusion that teacher assessment practices affected greatly student perception of the assessment. In terms of student achievement, better teacher-student classroom communication had no real impact. However, it resulted in a more positive student perception of the classroom practices. Dayal and Lingham (2015) reported on the varying conceptual understandings and perceptions viewed by teachers regarding assessment. These included assessment for improving teaching and learning, accountability for students and institution and also that assessment should be shunned due to its irrelevancy and invalidity, that is, that assessment results in negative outcomes and therefore should not be included. Teacher perception is extremely crucial as it will impact on teacher practice (Dayal & Lingham 2015). Dayal and Lingham's study consisted of researching Fijian teachers' conceptions of assessment by interviewing over seventy teachers. The interviews revealed that most teachers believed assessment was for formative purposes while these teachers generally viewed assessment as being primarily for learning purposes. For them, assessment's role had a learning purpose.

How assessment is perceived relates to whether one perceives it as being valid and reliable. Assessment can be the cause of regret among both teachers and students and both can hold the view that sometimes the assessment implemented was inequitable (Alm & Colnerud 2015). This study considered teachers' view of grading which was believed to be unfair. Many student questionnaires contain questions concerned with how students perceive assessment and the extent to which they perceive assessment as being a fair measure of their work (Dhindsa, Omar & Waldrip 2007; Dorman & Knightley 2006; Dorman, Waldrip & Fisher 2008). Discussion on validity and reliability results in perceiving assessment as a tool to gauge the measurement of something which it is supposed to measure and as being an accurate measurement of some element. In this case, measurement is in the educational context of an educational outcome

2.1. Reliability of Assessment

Reliability in assessment denotes the extent to which a particular test produces the same results (Carmines & Zeller 1979; Joughin 2010) therefore in this way causing it to be a consistent measurement of something. In school assessment, assessment can be a high stakes procedure leading, amongst other things, to students being accepted into university, Here, one hopes that the most capable students will consistently achieve the highest possible scores in these assessments so that the most talented and capable students will be chosen for university entry. Masters (2013) reiterates equating reliability in assessment with precision. This relates to the amount of confidence that can be ascribed to the conclusions resulting from educational assessments. Accurate ranking of students' learning progress is essential to reliability of assessment. Firstly, to be considered is why the particular assessments are being implemented and then how precise the results from these assessment methods are. If assessment measures students' literacy levels, the results from that assessment must accurately measure precisely that – students' levels of literacy.

The level of confidence in what is being demonstrated by the assessments must be high and levels of uncertainty in what the assessment shows must be minimal if ensuing conclusions on assessment are to be considered as reliable. Assessments making decisions on a student's mastery of knowledge must be adequate in order for the assessment to competently demonstrate a student's level of mastery of that particular knowledge. That is, assessments must be a true reflection of what they are being used for and therefore assessment tasks must be appropriately conducive to accurately achieving precise measurement (Masters 2013). Tasks must not be too easy or difficult for students if the tasks are to provide accurate proof of a certain element of students' achievement.

2.2. Validity of Assessment

The terms reliability and validity are often used conjunctively (Merriam 1998; Golfashami 2003; Miller et al. 2009; Joughin 2010) in particular in education. School assessments are ideally expected to be both valid and reliable and the processing of these assessment tasks is expected to be consistent.

Validity in assessment is about the extent to which a procedure or assessment task measures what it is intended to measure (Carmines & Zeller 1979). This means ensuring that appropriate measurement procedures are used for what is to be measured. It is quite possible for a measuring procedure to be highly reliable but not actually measure what it is intended to measure. Different procedures are utilised depending on what is intended to be measured. **Considering again the example of school assessments, the results of which influence, among other elements, high stakes results such as students' entry into university**, one is concerned with gauging the performance of students in mastering necessary skills and knowledge and not with, for example, ascertaining the shade of students' hair colour or their weight or height. Assessments must be purpose appropriate, that is assessments must be appropriate to the purposes for which they are being used. Appropriacy and relevancy must be applicable when selecting assessment tasks.

There are different types of validity. Construct validity (McGaw 2006) supersedes (Cronbach and Meehl according to McGaw 2006) the concept of the relation of measures or the measures to which assessments are related. An example of this is the one of school leaving or matriculation year examinations and entry to university. In the case of both Croatia and Australia, students' academic achievement is measured by a combination of external examinations and/or teacher assessment of assessment tasks. The ensuing results are usually illustrated in the form of grades which determine student acceptance into university entry. Construct validity relates to the basic theoretical construct that the purpose of assessment is supposed to be one of measurement.

Assessments across different subjects can vary because they are measuring different basic constructs. What is being assessed and how are vital. This can vary considerably depending on the subject. For example, in the subjects of music or sport, assessment tasks can demand evaluation of skill in physical activity. How well the skill is executed determines the final grade evaluation. In the subject of English, assessment can entail writing a 'for' or 'against' essay requiring that arguments

for and against a particular issue are included. How convincing these arguments are and how well they are expressed will determine student success. Answering a question in a foreign language measures grammatical accuracy to determine the level of success. How to use and apply mathematical formulas determines success in resolving equations. The assessment tasks are carefully chosen and not arbitrary in order to produce valid information about achievement in that particular subject. Another related issue is consequential validity (Messick 1996) where testing encourages students to focus on and learn what is going to be tested making only what is going to be test relevant for assessment (Sambell et al. 1997). It is concerned with the consequences of the outcomes of the assessment. This is where assessments and teaching to the exam become commensurate with each other. With high stakes assessment, teaching and assessment become focussed on this rather than on learning in a wider sense or education in general (Messick 1989 cited in McGaw 2006). For example, this can most often be observed with students who are final year students preparing for compulsory matriculation examinations for university entry and assessments tend to be mainly a reflection of the high stakes assessment or matriculation examinations. Students just focus on and learn what is going to be in the exam. This is because here the particular assessments such as external examinations have critical high stakes consequences such as determining students' future path in life and even their status in society. Certain vital outcomes occur as a result of a particular assessment. A sad example of this is in the matriculation examination for English taught as a foreign language in Croatian schools. Learning a foreign language encompasses the four skills of reading, writing, listening and speaking. However, as speaking is not tested in the matriculation 'matura' examination for English in Croatia which contributes to the danger of this vital skill being practised less in favour of the other three language skills which are contained in the final examination for English.

Reliability refers to how precise the conclusions about learner progress are (Masters 2015). Masters states that the more precise the conclusions are, the more confidence one can have in those conclusions and the less the uncertainty in assessment measurements is. In other words, has the assessment task resulted in the most competent or most highly skilled students been awarded the highest grades which will determine future success paths? Confidence is related to the amount of accurate evidence provided. Assessment tasks must be appropriate in order for tasks to provide precise results.

3. Research Methodology

The aim of this research was to estimate students' perception of fairness in teacher assessment of student achievement. As student perception of teacher assessment has both short term and long-term effects on various personal, educational, and social outcomes, it was hoped that a clearer estimate of student perception of teacher assessment would result in greater understanding of this concept and hence define more clearly steps that needed to be taken to achieve better outcomes in the wide variety of these critically important areas. This research was undertaken on the premise that the consistency with which assessment criteria are applied leads

to varying outcomes and impacts on students' perceptions of fairness. The research was aimed at examining assessment practices in Croatian secondary schools, how consistently teachers apply assessment criteria and students' perceptions of these practices. Investigating student perceptions of teacher assessments has never before been undertaken in Croatia; therefore, this study creates new knowledge and provides deeper understanding of the issue.

The research design here was intended to be both empirical in that students' perceptions were measured and quasi ethnographic in that it sought to interpret these perceptions through an understanding of the context. Student questionnaires, teacher interviews and analysis of teachers' assessment documents were used to collect data that were then analysed within case studies representing a teacher with two classes of students studying a particular subject. Classes of secondary school students from two secondary schools in Croatia studying the subjects of biology, Croatian or English as a foreign language formed the sample groups. These students were asked to complete a questionnaire that included Likert response type items based on the Student Perception of Assessment Questionnaire (SPAQ) and open-ended questions based on a questionnaire developed at the Centre for Schooling and Learning Technologies (CSaLT). The SPAQ items had five scales. The teachers of these students were interviewed and were asked to provide documents that illustrate how they assessed students in these three subjects.

3.1. Target Subject Areas

The target subjects consisted of a science subject which was biology and a humanities subject which was Croatian. These subjects were chosen intentionally as an example of a subject from two spectrums – one from science and one from humanities which could possibly result in enlightening comparisons of the results ensuing from the subjects originating from differing educational areas, namely, the sciences and the humanities. The third subject was be English as a Foreign Language which was a subject of particular interest to us the authors being teachers of English as a Foreign Language. Overall, the subject sample provided information from a range of disciplines that presumably entailed a variety of assessment tasks and assessment criteria and many variations in general.

Two secondary schools in Croatia were chosen. Secondary schools in Croatia are specifically focussed on different areas, Here, a school focussing on foreign languages and a school focussing on mathematics were chosen. The samples involved two schools to increase the potential to analyse different assessment practices being demonstrated. Two secondary schools were chosen with these two schools having a similar curriculum and therefore both had the three subjects of Biology, Croatian and English, which were the focus subjects of this research. Within these two schools, the biology class, Croatian class and English class were targeted together with their respective teachers. Also, final year secondary school students who were eighteen years of age were targeted. These students had much experience with teacher assessment after almost twelve years of school education and probably would be the most mature students in the school. Also, the students being eighteen years of age avoided having to seek and obtain parental consent to take part in the research.

The selection of teachers were 2 teachers who each taught those three selected subjects in that particular school and, of course, who were willing to participate. Consequently, final year classes of that respective subject in which those teachers teach were involved. The same students from each form class of year twelve students took the three subjects of Biology, Croatian and English as these are compulsory parts of the curriculum of these two schools. There were approximately 30 students per class, making a total of approximately 150 students. For two schools, this included approximately 330 students. Focussing on the three nominated subjects of Biology, Croatian, and English, two classes in each subject area for each individual teacher at two secondary schools were involved. Consequently, there were twelve classes all together. Two teachers were involved for each subject making a total of 6 teachers.

3.2 Student Questionnaire

The questionnaire was aimed at the students of the selected subjects. The students in each class taking one of the three subjects at each school completed a questionnaire. As the students were eighteen years of age accepting to fill out the questionnaire fulfilled consent requirements. The student questionnaire aimed to gain an insight into students' perceptions of teacher assessment in that subject. They were based on student reactions to teacher assessment in that particular subject. Again, asking the *right* questions (Chaboyer 2004) which results in the most comprehensive production of relevant information regarding the fairness and consistency of teacher assessment in these subjects was crucial.

Entire classes were questioned so highly structured closed questions were used to make these easy to answer for students. Types of questions were dichotomous or yes/no questions and rating questions using Likert response types from always to never. However as with teacher interviews, there were some open-ended response questions to encourage free student response to gain as much relevant information as participants were willing to give. The questionnaire was based on two existing instruments. Firstly, the SPAQ developed by Dorman, Waldrip and Fisher (2008) and secondly the student questionnaire developed by the Centre for Schooling and Learning Technologies (CSaLT) for the Digital Forms of Assessment project (Newhouse 2013). The SPAQ has five scales labelled *Congruence with planned learning, Authenticity, Student consultation, Transparency and Diversity*. The CSaLT instrument provided some open-ended questions and some questions related to the experience of students in completing assessments.

The SPAQ questionnaire was also used by Dhindsa, Omar and Waldrip (2007) to investigate secondary school students' perceptions of assessment in Brunei. In that particular research, based on gender and grade levels, results were comparable. However, differences were noted when comparing ethnic groups calling perhaps for change in teaching approaches and definitely warranting further research. Validity and reliability coefficients confirmed that the SPAQ instrument was appropriate for assessing students' perceptions in five assessment dimensions and hence suitable for use with students in Croatia.

The questions of this student questionnaire were classified according to five scales briefly described according to the following placed in Table 1.

Table 1.*Scale and Question numbers, Scale Description and Sample Question for each Scale.*

Scale and question numbers
1. Congruence with planned learning – questions 1-7
2. Authenticity – questions 8-14
3. Student consultation – 15-21
4. Transparency – 22-28
5. Diversity questions – 29-35

Scale description
1. Congruency – Extent to which tasks align with learning programme goals
2. Authenticity – Extent to which tasks reflect learner relevant real-life situations
3. Student consultation – Extent to which students are consulted about tasks employed
4. Transparency – Extent to which tasks are clear to learner
5. Diversity – Extent to which all students have equal chance to complete tasks

Sample question to describe scale
1. Congruency – My assessment is a fair indication of my work
2. Authenticity – I have answered questions on topics that has been covered in class
3. Student consultation – I am clear about the forms of assessment being used
4. Transparency – I know what is needed to successfully accomplish an assessment task
5. Diversity – I do work that is different from other students' work

(Adapted from Dorman, J. P., & Knightley, W. M. (2006). Initial use of the perceptions of assessment tasks inventory (PATI) in English secondary schools. *Alberta Journal of Educational Research,* 52, 3, 196-199).

3.3 Data Analysis for Teachers of English (Teacher 1 and 4)

As already mentioned, the subject of English taught as a foreign language in Croatia was of special interest to us the authors both being and working as teachers of English in Croatia. It was hoped that information gleaned from research in this subject would highlight areas in need of attention in our specific area of work. Therefore, we decided to focus on teacher 1 and 4 who were both teachers of English, but each in one of the two different schools selected for this research.

Each teacher selected for the study was interviewed individually about his/her assessment practices. Interview questions have to unearth information about teacher assessment practices while trying simultaneously to elicit the truth and reduce the possible effects of suspicion and lack of commitment (Barbour & Shostak 2005). Teachers were questioned regarding assessment tasks, methods of scoring, communication of expectations and results to students, strategies used to enhance fairness and consistency.

Interviews of teacher participant had to be audio recorded. Making extensive field notes while people are talking could result in loss of information by the gatherer. For example, by the time you write something down, you might have missed the next piece of information. It could also distract the interviewees' train of thought and can result in them forgetting what they wanted to say next. Hence, interviews were recorded to avoid distraction and hindering flow of information, simultaneously enhancing accuracy and amount of information given.

3.3.1 Quantitative and Qualitative Data for Teacher 1

From high school 1, this teacher named Teacher 1 had two classes of English as a foreign language: class 1, numbering 19 students of who 14 are female and 6 are male, and class 6 numbering 29 students of whom 22 are female and 7 are male. The results for these two classes using SPSS are shown in Table 2.

Table 2. Results from student responses to questionnaire items for Teacher 1.

Teacher		N	Mean	St. Dev.
Congruence1	1	25	3.6800	.34077
	6	29	3.5764	.39583
	Total	54	3.6243	.37159
Authenticity1	1	25	3.3943	.27130
	6	29	3.4483	.46562
	Total	54	3.4233	.38549
Consultation1	1	25	2.4457	.36728
	6	29	2.3596	.30213
	Total	54	2.3995	.33345
Transparency1	1	25	3.4114	.38975
	6	29	3.4828	.49841
	Total	54	3,4497	.44868
Diversity1	1	25	2.3771	.37544
	6	29	1.9113	.27109
	Total	54	2.1270	.39701

A difference in the means between classes concerning Diversity was indicated. The means for all other areas is very high, well over 2.3. Table 3 provides the ANOVA differences between the two groups of Teacher 1.

Table 3. ANOVA differences between the two groups of Teacher 1.

Congruence1	Between Groups	.311
	Within Groups	
	Total	
Authenticity1	Between Groups	.612
	Within Groups	
	Total	
Consultation1	Between Groups	.349
	Within Groups	
	Total	

Transparency1	Between Groups	.565
	Within Groups	
	Total	
Diversity1	Between Groups	.000
	Within Groups	
	Total	

As previously mentioned, there is significant difference between these classes concerning Diversity, that s equal chances for successful completion of work indicating that in this area the teacher is more flexible in approach.

Qualitative data from two teachers of English will now be presented. From high school 1, this teacher named teacher 1had two classes of English as a foreign language: class 1, numbering 19 students (labelled A) of who 14 are female and 6 are male, and class 6 (labelled B) numbering 29 students of whom 22 are female and 7 are male. Following the teacher's answers to the structured interview are student open ended comments from the student questionnaire.

1. What assessment tools do you use?
1. This teacher used oral and written tasks and repeated the fact that she used oral and written tests. She also used essay and grammar tasks. She attempted at all times to assess all 4 foreign language skills of reading, writing, listening and speaking.

2. How often do you use these assessment tools?
2. This teacher assessed three times per student per year oral examining plus written tests 4-5 times per year. This is obligatory. There are 2–3 assessments per semester. Essay composition is 2 times per year per student 2 grammar 2 vocabulary tests per year oral twice a year.

3. Do you give the students assessment criteria?
3. This teacher stated she explained really in fact what she expected from students so they are clear. Yes, the teacher repeated continuously so that students were clear on what she expected.

4. If so, in what form?
4. The teacher told them what the criteria were, and the criteria were written out. The teacher believed the students should have the assessment criteria. The criteria explain what and how marks are allocated in a certain way and what they mean. The school has them.[2] She claimed she always told them to write them down, The teacher claimed students copied down the assessment grids done by the ministry. At county level, the teacher claimed they had to decide on the criteria for assessment and have them written up and in the schools and available to all.

---

2   All schools by Croatian Education Department decree, wrote up the assessment criteria for their subject at a regional level. All teachers of that particular subject were assembled and wrote up assessment criteria for that subject in that school. These criteria are expected to be in the school available to all relevant persons, students, teachers, parents and researchers to see upon request. The authors have a copy of these criteria.

5. How consistently do you apply them?

  5. This teacher repeated that she applied the criteria very consistently, very consistently in answer to this question, but did admit that she does not always do so not in full, and that it depended upon the task. Sometimes a written task for the same student can be a 4 sometimes a 2.

6. What do you base your assessment most on?

  6. This teacher stated that thoroughness, hardworking, diligence, and not being superficial were the most important when assessing. Apart from that, accuracy and range were important in assessing a foreign language.

7. What do you think students think about how fair your assessment is?

  7. This teacher believed the first impression students had of her is that she is too strict, but then they get used to her and her assessment and the students were satisfied because they know what she as the teacher expects. This teacher believed that some students think that she is not too strict. She believed students get that impression first from her, but get out of it quickly because of the teacher's nature which she believes to be very compassionate. This teacher said it is not in her nature to be too strict.

Now follow students' comments when asked to voluntarily provide their thoughts on assessment in this subject which is the subject of English. The first comments, A, signify the first of the classes taught by Teacher 1 who is an English teacher in one of the schools. The second comments, B, signify the second class of English also taught by Teacher 1 in Secondary school 1 of the two chosen secondary schools for the purposes of this research.

A.1.
One student said assessment was objective as a positive comment and also described it as educational and fun
One said sometimes they cannot answer questions objectively and tasks are not suitable for age of students.
One said they were learning new things.
One said assessment checks knowledge.
One said assessment was stressful.

B. 6.
Seven said assessment reflected real life application and was useful.
Four liked reading the books in English.
Two said there was good communication with teacher.
Six students liked English literature.
Three said that oral tests were arranged in advance.
Four said they liked doing the tasks that improved vocabulary and grammar.
Two said assessment was fair.
One said assessment was objective.
One said to get an A you must get As all year.
One said they were examined on things they learned in class.

One said assessment was not stressful.
Two said tests were suitable and not demanding.
One said subject had good preparation for matriculation.
One liked that they could express their opinion.
One liked text analysis.
One did not like analysing extra texts.
One liked 5reading tasks.
One did not like reading books in English.
One said content was irrelevant.
Two said tasks were too easy and too boring.
Five said textbook and essay topics were boring.
Two said they didn't like surprise tests.
Two believed listening tasks were not true indicators of knowledge.
One said some tasks had two meanings.
Other comments noted they liked the teacher who was involved.
Two preferred oral examinations.
One said good preparation for tests.
One said they didn't like preparing for the matriculation.
One said this was their favourite subject- they liked learning and practising English.
One said textbook was not used much.
One said content was important.
One said tasks improved logical thinking.
One liked that they learned quickly.

The only negative comment from three students in one class was that the topics/textbooks were boring. One comment was irrelevance of content. The rest of the comments were overwhelmingly positive (15) regarding assessment and teaching content. There are comments about teacher involvement, notification of assessment and students expressing their opinion so participation with teacher. This correlated with the means of most scales which are also high.

From the teacher, she admitted there was some inconsistency in her assessment, but despite this, the students' comments seemed to indicate students were satisfied with the assessment. In fact, there was a strong indication of student satisfaction with the teacher the consequence appears to be student overall satisfaction with everything related to this class of English. There seemed to be some connection between student satisfaction with teacher and student perception of the subject, of which assessment was an integral part of the research carried out.

3.3.2 Quantitative and Qualitative Data from Teacher 4

Teacher 4 taught two classes of English in the second of the two chosen high schools. These were named class 7 and class 12 for the purposes of this research. These two classes, labelled class G and class H for the purposes of indicating student responses to fairness teacher assessment, the fundamental of this research, each numbered twenty-five students making a total of fifty students in these two classes.

The numerical student responses from the Likert scale-based questions from the

questionnaire given to students were appropriately analysed on teacher 4's two targeted classes numbered 7 and 12 constituting the five scales of Congruence, Authenticity, Consultation, Transparency and Diversity through use of the SPSS statistical programme for social sciences (see Table 4).

Table 4. Results from student responses to questionnaire items for Teacher 4.

| Teacher Four		N	Mean	st. Dev.
Congruence1	7	29	3.2808	.56305
	12	25	3.4914	.30045
	Total	54	3.3783	.46862
Authenticity1	7	29	3.0542	.58206
	12	25	2.9200	.49501
	Total	54	2.9921	.54268
Consultation1	7	29	2.5616	.42675
	12	25	2.2286	.30861
	Total	54	2.4074	.40918
Transparency1	7	29	3.3498	.49130
	12	25	3.1943	.33977
	Total	54	3.2778	.43118
Diversity1	7	29	2.4187	.32156
	12	25	1.8229	.43307
	Total	54	2.1429	.47906

The items were then processed using ANOVA which gained the following information for this teacher and their two classes (see Table 5).

Table 5. ANOVA differences between the two groups of Teacher 4.

Congruence1	Between Groups	.100
	Within Groups	
	Total	
Authenticity1	Between Groups	.370
	Within Groups	
	Total	
Consultation1	Between Groups	.002
	Within Groups	
	Total	
Transparency1	Between Groups	.189
	Within Groups	
	Total	

	Between Groups	.000
Diversity1	Within Groups	
	Total	

A significant difference between the two groups, which is indicated by a result of less than 0.05, is noted in the areas of student consultation and diversity. Therefore, students in these two classes had significant differences regarding the consultation by teachers with them on assessment and the opportunities provided to them for completing assessment tasks. This fact is, although the classes are grouped homogenously, the teacher is the same for two different classes of the same subject at the same school, it can be concluded that nevertheless this teacher seemed to have affected a significant difference from students' perspectives in approach to assessment practices in the area of Consultation and Diversity. Diversity is particularly crucial because this research focusses on fairness in teacher practices of student assessment and Diversity deals with the extent to which all students have equal chances to complete assessment tasks. Various reasons are possible. This could be the result of the teachers behaving differently to different classes regarding assessment. Interestingly there is no statistically significant difference where concerning Congruency, the extent to which tasks align with what is taught, is dealt with. Therefore, clarity and relevance of tasks seemed to be agreed upon by these students, but there appears to be significant indication of unfairness in opportunities to achieve task success. This was reflected in teacher and student comments.

Comparing all teachers involved in this research, regarding Diversity, the results are depicted in Figure 1.
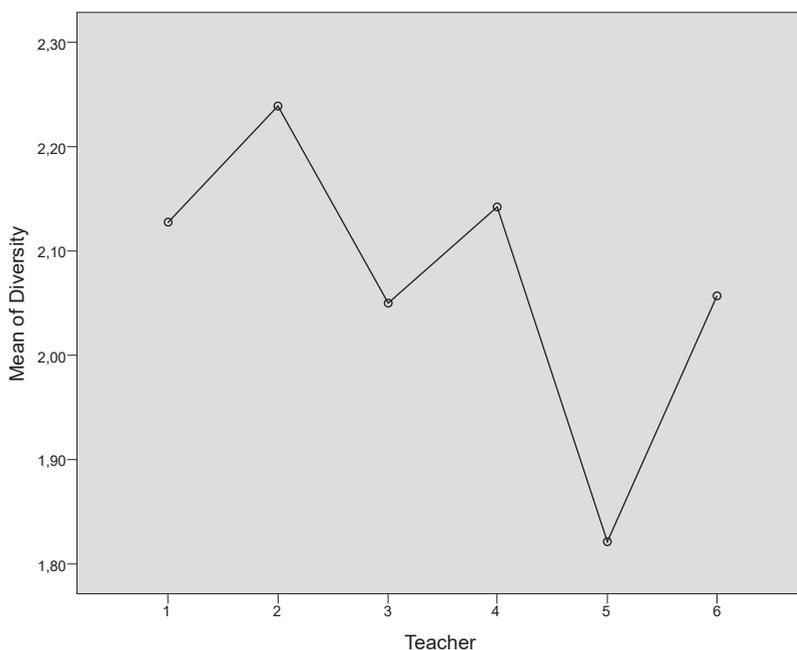
Figure 1. Mean of diversity according to teacher.

We can see here that this teacher 4 mean for Diversity is higher than that for teachers 5, 6 and 3, but Teachers 1 and 2 were higher.

Teacher four's responses to the interview questions are summarised as follows. Also added are the teacher's students' open-ended responses to the negative and positive aspects of assessment and assessment in general. The teacher's students' comments are divided into two groups, G7 for class seven and H12 for class 12, both signifying classes taught by this teacher of English from the second secondary school. Student comments are placed under the related teacher interview question, if there were any.

1. What assessment tools do you use?
This teacher believed that the students had too much assessment and too often. She said students were given 8 tests per student per year plus oral grades.

> G. 7.
> A student believes that doing listening tasks is a problem because s/he rarely has enough time for completing the test.

> H.12.
> A student thinks that sometimes the tasks don't make sense.

2. How often do you use these assessment tools?
This teacher gave them two grades one for spoken language and one for vocab reading, listening. She claimed to use written tasks such as essay, compositions, grammar tests, when the end of a grammar unit or activity is reached. She claimed

she placed special emphasis on the tenses. In oral tests it was difficult because you have to grade on the spot. Oral can be projects, persuasive speech, research. This teacher awarded two grades which are for speaking skills and understanding respectively.

> G. 7.
> Two students commented that assessment tools included essays and oral examination.

3. Do you give the students assessment criteria?

This teacher claimed she talked about what was going to be assessed, what she is looking for. This teacher stated that she and the class also talked about assessment at the end to explain what they have got. The teacher admitted that she did not have a set of assessment criteria written down.

4. If so, in what form?

This teacher claimed that before she graded, she gave students criteria both in oral and written form, the teacher told them to write it down and what the grade was for each criterion. Then the teacher admitted that she actually really just told them because this teacher had the assessment criteria in her head and that she just knew how to assess students. It depended on the task according to this teacher..

It should be noted here that, although this teacher did not have any/did not know about any official criteria, the school actually provided me with assessment criteria. One of the teachers in charge stated that all schools were required to produce assessment criteria. The teachers had to reach consensus on these criteria and had to have them in the school. Each teacher was to apply these criteria for their respective subjects. These criteria were there for any parents or even students for that matter who wished to see them.

> G.7.
> A student commented that this teacher had a strange grading system.
> Another student said that before tasks for vocabulary and grammar, they prepared in class.

5. How consistently do you apply them?

This teacher was of the opinion that strict was fair. She claimed to adhere to criteria as consistently as possible. This teacher also gave written comments as well. The most difficult part for this teacher was how to assess creativity. Actually, she said that creativity was not in the criteria, but adhered to criteria and took into consideration the age and topic and what the teacher believed creativity to consist of. All students had the same topic. How this teacher assessed all depended on the task or topic.

> G. 7.
> One student said that assessment was not good enough.
> One student said that if you fail a test you can do it again.

One student said that if she or he did the test well she or he got a good mark; since the teacher explained everything in class and that is was difficult to get a bad mark even if you didn't study.
One student said that before tasks for vocabulary and grammar, they prepared in class and that if you're missing a point for a better mark the teacher will give it to you.
Another said that if you fail a test you can do it again.

H. 12.
One student stated that sometimes their mark wasn't equal to what they actually knew and that sometimes tests were too difficult.
Another said that assessment on compositions was unfair.

6. What do you base your assessment most on?

The teacher said it depended on the skill or tasks being assessed, and that then there was the issue of students who make an effort, and what to do with that. Where does effort get included was the concern of this teacher. She said it was hard to get a 1 (F) in this subject.

7. What do you think students think about how fair your assessment is?

The teacher thought that there were too many grades, too much being tested all the time. Namely, there was grammar, knowledge, everything. She thought that students did not mind assessment and that it just depended on the student. That is, some found it stressful, some just rote learn, some like it, some liked getting grades.

G. 7.
One student stated that assessment just checks knowledge from the book and not from the real world.
One said he or she will learn English better.

H.12
One student said that they will know English better.
One said it checked whether they had learnt the content and that it encouraged them to learn.
One said that they learned how to interact in English
One stated that sometimes tests are too difficult.

The teacher and students' comments seemed to indicate that assessment criteria were not negotiated or known in advance by either the teacher or students and definitely not at the same time. Consequently, assessment was based on teacher impression and what seemed important to the teacher at the time of marking and depending on the teacher's perspective of what was important for that task. Criteria did not appear to be applied consistently which was reflected in the sometimes-contradictory nature of the teacher responses. Students seemed to be assessed on what was relevant to the subject of English and the comments seemed to indicate that they were learning how to be better at English.

4. Research Results and Variance on Scale Scores by Subject

Given that obvious differences existed between teachers in terms of students' responses to questionnaire items, analysis of differences between subjects was warranted. Results from an ANOVA were presented in Table 2. There was a statistically significant difference between subjects for Authenticity ($p < 0.001$), Consultation ($p < 0.001$), Transparency ($p < 0.05$) and Diversity ($p < 0.001$), but not for Congruence. For the reliable scales of Congruence, Authenticity and Transparency, the subject of English had the highest mean scores. On the other hand, Biology and Croatian scored the lowest in the five scales according to student responses. For example, Croatian scored the lowest in Authenticity and Biology the lowest in Congruence, which were both reliable scales. These differences can be best seen in the graph presented in Figure 2 and the mean scores listed in Table 6. Here the scales are presented in different colours and student responses to English have evidently generally gained the highest mean scores.

Table 6. *Variance on mean scores for scales by subject.*

		N	Mean	Std. Dev	Sig
Congruence	Biology	106	3.39	.43	
	Croatian	116	3.47	.40	0.135
	English	108	3.50	.44	
Authenticity	Biology	106	3.02	.43	
	Croatian	116	2.98	.39	<0.001
	English	108	3.21	.52	
Consultation	Biology	106	2.25	.34	
	Croatian	116	2.40	.27	<0.001
	English	108	2.40	.37	
Transparency	Biology	106	3.20	.39	
	Croatian	116	3.22	.50	0.015
	English	108	3.36	.45	
Diversity	Biology	106	1.92	.38	
	Croatian	116	2.15	.43	0.001
	English	108	2.13	.44	
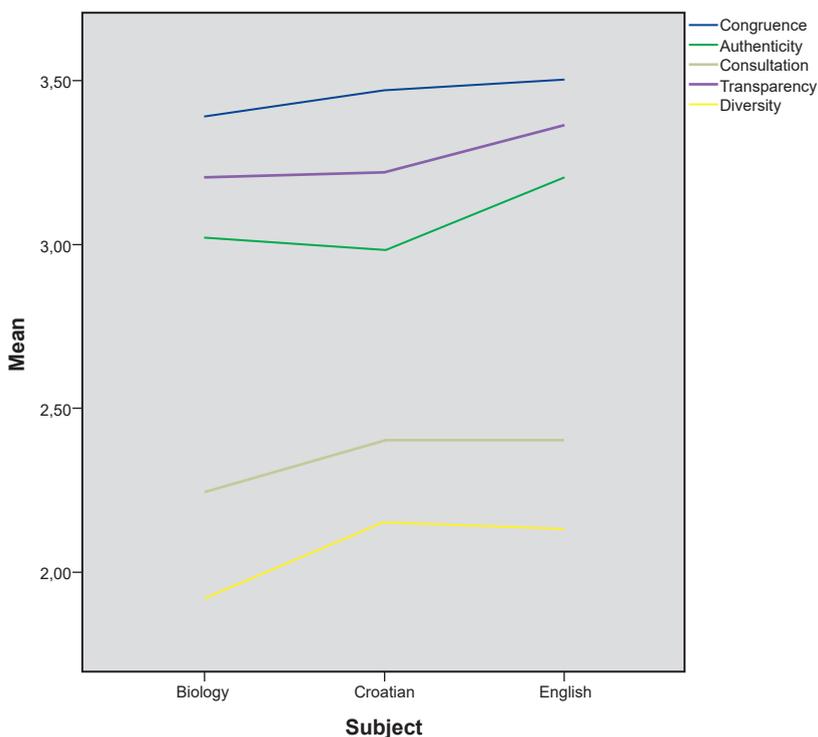
Figure 2. Mean responses according to subject and scale in graphical form.

Table 7. Mean scores for all students by scale and subject area.

Scale	Biology	Croatian	English
Congruence	3.39	3.47	3.50
Authenticity	3.02	2.98	3.03
Consultation	2.24	2.40	2.40
Transparency	3.20	3.22	3.36
Diversity	1.92	2.15	2.14

The mean on the Congruence scale was highest for all three subjects (see Figure 2), and highest for the subject of English (see Table 7) seeming to indicate that students in all subjects tended to perceive that assessment tasks were commensurate with what was being taught in class, particularly in the subject of English. Biology scored the lowest with a mean lower than 3.40. Again, for the Authenticity scale, the subject of English can be said to be perceived by students as most having assessment tasks that reflect real life. The score was above 3.20 for English compared to less than 3.05 for Biology and even below 3.00 for Croatian. For the Consultation scale the mean scores were the same for English and Croatian indicating that on average students

perceived the same amount of teacher consultation on assessment tasks. They perceived slightly less consultation in Biology with a mean of 2.24.

The subject of English had the highest mean for the Transparency scale indicating a perception that the assessment tasks in the subject of English tended to be clearer than in the other subjects. It was assessed by the mean student response, which was well over 3.35 compared to Biology and Croatian which were below 3.25. For the Diversity scale, the subject of Croatian scored the highest, over a mean of 2.15, although this scale was proved to be unreliable by Cronbach's Alpha.

Overall, regarding subjects, the subject of English scored highest for all scales apart from the unreliable scales of Diversity where Croatian scored the highest, and of Consultation English scored equally with the subject of Croatian. Both Consultation and Diversity were not proved to be reliable or statistically significant. Therefore, in terms of scales that were reliable, the subject of English scored the highest, that is, that the student sample was most satisfied with the subject of English in terms of the reliable student perceptions of assessment in the three analysed subjects. The teacher scoring the highest in the reliable scales was the Teacher 1who taught English indicating the strongest trend of student positivity of student perception of satisfaction in the subject of English and to the teacher of English.

It needs to be noted that concrete research results in teacher fairness in student assessment was severely lacking in the literature. Hence the idea of this research was concepted to compensate for this deficiency.

5. Conclusion

The main issue arising from the research results and the basis for further investigation, we feel, is why students perceive the subject of English the most positively. As teachers of English this question could implicate English language teaching and assessment methodology, teacher training and other relevant elements of teaching and learning.

The limitations of the research must be emphasised in that only two secondary schools were involved and only six teachers from each school and only three subjects. The research sample would have to be increased to ensure greater validity and should include vocational schools as well as grammar schools. Further research would be needed as this is a topic extremely relevant at this time of reform in educational policy in Croatia resulting from Croatia's entry into the European Union.

References

Alkharusi, H. 2014. "Classroom Assessment Communication, Perceived Assessment Environment and Academic Achievement: A Path Analysis." *British Journal of Education,* 8, 2, 117–126.

Alkharusi, H. 2015. "An Evaluation of the Measurement of Perceived Classroom Assessment Environment." *International Journal of Instruction,* 8, 2, 45–54.

Alkharusi, H. 2008. "Effects of Classroom Assessment Practices on Students' Achievement Goals." *Educational Assessment,* 13, 4, 243–266.

Alm, F. & Colnerud, G. 2015. "Teachers' Experiences of Unfair Grading." *Educational Assessment,* 20, 2, 132–150.

Blazevic, T. 2019. *An Investigation of Secondary Student Perceptions of Fairness about Assessment Processes Implemented by their Teachers.* Doctoral thesis. Joondalup: Edith Cowan University.

Bruner, J. S. 1960. *The Process of Education.* Cambridge: Harvard University Press.

Carmines, E.G. & Zeller, R. A. 1979. *Quantitative Applications in the Social Sciences: Reliability and Validity Assessments.* Thousand Oaks: Sage Publications Ltd.

Dayal, H. C. & Lingam, G. I. 2015. "Fijian Teachers' Conceptions of Assessment." *Australian Journal of Teacher Education,* 40*,* 8, 43–58.

Dhindsa, H. & Omar, K. & Waldrip, B. 2007. "Upper Secondary Bruneian Science Students' Perceptions of Assessment." *International Journal of Science Education*, 29, 10, 1261–1280.

Dorman, J. P., Waldrip, B. G. & Fisher, D.F. 2008. "Using the Student Perceptions of Assessment Questionnaire (SPAQ) to Develop an Assessment Typology for Science Classes." *Journal of Science Education*, 1, 9, 13–17.

Dorman, J. & Knightley, W. 2006. "Development and Validation of an Instrument to Assess Secondary School Students' Perceptions of Assessment Tasks." *Educational Studies,* 32, 1, 47–58.

Golfashami, N. 2003. "Understanding Reliability and Validity in Qualitative Research." *The Qualitative Report,* 8, 4, 597–606.

Joughin, G. 2010. *A Short Guide to Oral Assessment.* Leeds: Leeds Metropolitan University; University of Wollongong. https://www.qub.ac.uk/directorates/AcademicStudentAffairs/CentreforEducationalDevelopment/FilestoreDONOTDELETE/Filetoupload,213702,en.pdf.

Masters, G. N. 2013. "Reforming Education Assessment, Imperatives, Principles and Challenges*." Australian Education Review*, 57, 9–68.

McGaw, B. 2006. "Assessment Fit for Purpose". International Association for Educational Assessment, 32ⁿᵈ Annual Assessment in an Era of Rapid Change: Innovations of and Best Practices, May 21–26 2006, Singapore.

Merriam, S.B. 1998. *Qualitative Research and Case Study Applications in Education, Revised and Expanded from "Case Study Research in Education."* San Francisco: Jossey-Bass Publishers.

Messick, S. 1996. "Validity and Washback in Language Testing." *Language Testing*, 13, 3, 241–256.

Miller, M.D., Linn, R.L. & Gronlund, N.E. 2009. *Measurement and Assessment in Teaching*. Upper Saddle River: Merrill/Pearson.

Newhouse, C.P. & Cooper, M.G. 2013. "Computer-based Oral Exams in Italian Language Studies." *Recall,* 23, 3, 321–339.

Newhouse, C.P. 2010. "Digital Forms of Performance Assessment." *In Proceedings of 21st Australiasian Computers in Education Conference* (pp. 8p.). Melbourne. Australian Council for Computers in Education. https://ro.ecu.edu.au/ecuworks/6611/.

Newhouse, C.P. 2013. "Computer-based Examinations in Schools: Freedom from the Limitations of Paper?" *Research and Practise in Technology Enhanced Learning,* 8, 3, 4311–447.

Sambell, K., McDowell, L. & Brown, S. (1997). "'But is it fair?': An Exploratory Study of Student Perceptions of the Consequential Validity of Assessment." *Studies in Educational Evaluation*, 23, 4, 349–371.

Appendices

List of tables

List of figures

UČENIČKE PREDODŽBE O VRJEDNOVANJU SREDNJOŠKOLSKIH NASTAVNIKA S NAGLASKOM NA PREDMET ENGLESKI JEZIK

Sažetak

Ovaj je rad usredotočen na učeničke predodžbe o pravdenosti vrjednovanja u trima srednjoškolskim predmetima na temelju provedenoga istraživanja koje ispituje utjecaj nastavničkoga ocjenjivanja na učeničke predodžbe o pravednosti ocjenjivanja u srednjoškolskome sustavu obrazovanja u Republici Hrvatskoj. Provedeno istraživanje ujedno je kvantitativno i kvalitativno. Dakle, ono je bilo namijenjeno kao empirijsko, mjereći učeničke predodžbe, i kvazietnografsko, interpretirajući ove predodžbe razumijevanjem konteksta. Učeničkim anketama, na temelju uzorka 330 učenika, intervjuima s nastavnicima, na temelju uzorka 12 nastavnika, i analizama nastavnikove dokumentacije za vrjednovanje koristilo se sa svrhom prikupljanja podataka, koji su kasnije analizirani unutar studija slučaja koje su predstavljale pojedinoga nastavnika s dvama razredima učenika koji uče određeni predmet. Dva srednjoškolska razreda koji uče predmete Biologija, Hrvatski jezik i Engleski jezik, kao strani jezik, predstavljaju uzorak skupine. Od ovih se učenika zatražilo ispunjavanje upitnika koji je sadržavao čestice odgovora po Likertovoj ljestvici na temelju *SPAQ*-a (Upitnik o učeničkim predodžbama o vrjednovanju) i pitanja otvorena tipa na temelju upitnika razvijena u *CSaLT*-u (Centar za obrazovne tehnologije i tehnologije učenja). *SPAQ* čestice doprinose pet ljestvica. Nastavnike navedenih učenika intervjuiralo se i zatražilo dokumentaciju koja pokazuje načine vrjednovanja učenika u ovim trima predmetima. Istraživanje je pokazalo razlike među nastavničkim odgovorima i predmetima te razlike među učeničkim odgovorima u različitim razredima istoga nastavnika, što vjerojatno upućuje na ulogu čimbenika nastavnikovih crta ličnosti. Ovo istraživanje također otkriva kako su među trima istraženim srednjoškolskim predmetima učeničke predodžbe o predmetu Engleski jezik procijenjene najpozitivnijima u uzorkovanoj skupini učenika.

Ključne riječi: dosljednost, engleski kao drugi jezik, pravednost, učeničke predodžbe, vrjednovanje