

Exploring the Effects of Genre-based Portfolio Assessment on EFL Writing With Focus on Learner Engagement

Natasha Pourdana (✉ natasha.qale@gmail.com)

Department of English Language Teaching and Translation, Islamic Azad University, Karaj Branch

Original article

Keywords: Genre, Learner Engagement, Portfolio Assessment, Writing

Posted Date: May 5th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-482242/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

With controversial implications in language learning context, there is not enough evidence for learners' engagement in genre-based portfolio assessment (PA) and its impacts on their writing improvement. To fill the gap, this case study examined 46 EFL undergraduate students' performance on descriptive and narrative writing tasks. In 12-week PA design, feedback points were collected from teacher formative assessment of the students' descriptive and narrative writing, with reference to the genre-specific indicators in West Virginia Department of Education descriptive writing rubric and Smarter Balanced narrative writing rubric, respectively. Statistical results reported the significant impact of PA on improving components of word choice and grammar, development and organization of ideas in student descriptive writing, with no sign of improvement in their performance on descriptive writing post-test. Statistics also supported the positive impact of PA on improving the components of elaboration of narrative, language and vocabulary, organization and convention in student narrative writing, and their performance on narrative writing post-test. The qualitative data on students' engagement in PA was collected from their reflective journals. After inductive content analysis, the students' self-reports were schematized, and their level of engagement was rendered in terms of their approval of usefulness and novelty of PA, frequent mismatch between student self-assessment and teacher written feedback both in quality and quantity, 'sensitivity' or focus of teacher feedback to some writing features over others, applicability of teacher feedback to revision process, and overall perception of writing improvement. The paper provided teaching implications for EFL practitioners and suggestions for future research.

1. Introduction

Since the early 1990s, assessment for learning (AfL) has reached worldwide audience for improving teaching and learning in educational context (Cowie, 2005; Klenowski, 2009; O'Shea, 2019). Several studies have examined how AfL might benefit second/foreign language (L2) teaching pedagogy, language learning, and L2 learner performance (Brown & Abeywickrama, 2010; Dann, 2014; Darling-Hammond & McCloskey, 2008; Earl, 2013). As Ramaprasad (1983) conceptualized, AfL in language classroom needs the L2 learners' perception of a gap between a long-term goal and their status quo, as well as their commitment to bridge the gap to attain the goal. Ideally, either language learners will engage in self-assessment to generate the information about the gap, or the teachers themselves will explore the gap and to provide feedback about it to the students. Ultimately, the action to close the gap will be taken by the fully engaged students in the process of learning (Sadler, 2010). But in reality, L2 teachers and learners have more critical steps to take. In AfL practice, the teacher needs to reinforce the capacity in the students to engage, to diligently discover the gap, and to take full responsibility for carrying out remedial actions. Thus, L2 learner self-engagement is not an option; it is a survival kit. However, the focus on learner engagement is not a common practice in most L2 classrooms, as the majority of language teachers do not welcome such shared responsibilities with students (Ecclestone, 2007; Hargreaves, 2005). In essence, as Black and Wiliam (2009) rightfully disputed, the practicality of AfL at language

learning classroom level has remained insufficient and more evidence is needed to support the real benefits of various types of AfL, including portfolio assessment.

Portfolio assessment (PA), as a common platform of AfL, largely demands L2 learners to actively engage in self-assessment and self-reflection (Lam, 2013; 2014) for reaching a closure in the learning gaps. This is advised through redrafting and writing reflective journals (Hamp-Lyons, 2016; Lam & Lee, 2010). Yet, the full practice of PA in L2 setting has faced massive problems, such as teachers' AfL malpractice (Harris & Brown, 2009; Willis, 2011) or lack of learner self-engagement (Lee & Coniam, 2013; Li, 2010). Therefore, Hyland and Hyland (2006) called for more research on the aspects of PA impact on learning of writing skills in English as a Foreign Language (EFL) settings.

Some research examined the teachers' experience with various models of PA (i.e., progress, workshop, and showcase) (Lam, 2017; 2018a; Lee, 2016; Lam & Lee, 2010), the effects of PA on boosting L2 learner autonomy and metacognitive awareness (Aydin, 2010; Carless, 2011; Hirvela & Sweetland, 2005), and learner text revision strategies (Hamp-Lyon & Condon, 2000). However, research findings on how teacher portfolio assessment might impact the L2 learner engagement in genre-based writing have remained unclear with limited empirical evidence (Hamp-Lyons, 2007). In the same vein, while the use of portfolio approach to collect student performance on different genres of writing has been well-reported in L1 writing (Hyland, 2004; 2007), yet the contribution of portfolio assessment into EFL learners' genre-based writing performance is largely under-documented (Lam, 2019; Wiliam, 2006). To void the gap in the literature, this study aimed to set a genre-based PA platform to investigate the role of teacher formative assessment in EFL learners' degree of engagement in descriptive and narrative writing progress.

1. Literature Review

A writing portfolio is a collection from a large body of students' written works, often endorsed with reflection pieces of writing by the students. As a popular platform for self-regulated learning and evaluation, portfolio assessment (PA) is often assumed a better-quality alternative to traditional, product-oriented assessment for improving the student writing performance and long-term learner engagement (Benson, 2006a; Hirvela & Sweetland, 2005; Li, 2010; Mohamadi, 2018). However, as Condon and Hamp-Lyons (1994) argued, "the portfolio has simply been accepted on faith, on writing specialists' feeling that the portfolio is better" (p. 277). Despite reported educational benefits, PA has remained controversial when utilized in classroom situations, namely due to L2 teacher inflexibility (Xu & Brown, 2016), insufficient and unwilling student engagement (Li, 2010), complicated and holistic grading system (Song & August, 2002), and lack of school support (Lam, 2018a; Lam & Lee, 2010).

As Lam (2018b) indicated, since writing portfolios are reported to sustain students' close attention to their own progress of writing, their active engagement in teacher feedback is central (Burner, 2014). Furthermore, for a powerful PA experience, L2 writing teachers should prime such student self-reflectiveness. On the other side of the isle, however; evidence of how well students comprehend and engage in working portfolios in L2 context is still anecdotal and under-researched (Hamp-Lyons, 2007; Hamp-Lyons & Condon, 2000). To reach confidence in the student engagement and self-assessment in

PA, L2 writing teachers may need 'scaffolding' the students in terms of tutorials on the entire portfolio process (Carless & Boud, 2018), using examples and prompts (Gregory, Cameron, & Davies, 2001), extending deadlines to sustain their engagement (Lam, 2014), and training them to writing assessment rubrics (Panadero & Romero, 2014). Romova and Andrew (2011) emphasized the critical role of self-assessment practice in PA as it warps student persistence, academic engagement (Finn & Zimmer, 2012), and ultimate achievement in PA. Successful engagement depends on *how well* L2 learners understand the goals of PA, *how soon* they picture the distance between their own status and the goals (Caner, 2010), and *what they actually do* to reach the goals. In other words, learner engagement is the bread and butter for effective learning, yet the topic has been overlooked in the mainstream research on PA (Price et al., 2010; Steen-Utheima & Hopfenbeck, 2018).

Apart from controversies over its aftermath in language learning context, PA is still assumed as a powerful pedagogical and assessment alternative, mostly because it reinforces the L2 learners' "understanding of writing as a socially-situated process in academic discourse communities" (Duff, 2010, p. 169). In doing so, genre-based writing PA can assess both microscopic (i.e, mechanical, formal) and macroscopic (i.e., textural, discursive) aspects in L2 learner writing progress (Borg, 2003). Learning writing genres such as narrative, descriptive or expository is one of the critical issues of all times in SLA research (De Fina & Georgakopoulou, 2015; Shober, 1996). Hyland (2003) celebrated engagement in the genre-based writing process as empowering, dialogic, and systematic metacognitive awareness in language learners. Adopting a genre-based approach, Hinkel (2002) also suggested that to develop effective written discourse, EFL students should master "the mechanical aspects of composing sentences, paragraphs, and larger units of discourse that correspond to the dominant genres of the academy, a specific field, or both" (p. 57). While the process approach has an eye on the L2 writers overflow of ideas, the genre-based approach has switched its focus to the socio-literacy of the L2 writers in generating real texts that properly address the target discourse community (Hyland, 2003). While Badger and White (2000) believed that the product, process, and genre-based approaches to writing interplay, Romova and Andrew (2011) argued that the genre-based approach only integrates with the process approach, by "adding focus on text/context, and emphasizing the role of language in written communication" (p. 114). Hamp-Lyons and Condon (2000) argued that writing PA can best fit with such 'genre-process nexus' approach.

Genre-based approach to analysis of written narrative is no longer the sole responsibility of literary studies. Narrative analysis has entered the realm of human sciences and professional practice, including psychology and learning L2 writing (Roohani & Taheri, 2015). Assumed as an art or gift of storytelling, narration is made through every minute of every day in our life, so that we make narration plenty of times (Abbott, 2002). Lou-Conlin (1998) defined written narrative as a system of gradual development through which the writer entertains with the logical sequence of ideas and events. Narration is mostly done with the purpose of maintaining the readers' interest in a given event or personal experience narrative (PEN) (Labov, 2001). In the same vein, descriptive genre of writing gives certain attributes to a person, place or chain of events in detail. Such entities should be described in such a way that the reader can capture the topic and enter the writer's experience. Descriptive writing is considered as a means to improve other

genres of writing such as narrative and expository or perhaps as a dominant strategy in writing academic texts (Birjandi & Hadidi Tamjid, 2012).

In line with genre-based approach to PA, EFL learners may have a chance to engage in gaining control over a variety of genre-based writing such as narrative and descriptive in the target discourse. However, literature on PA mostly pertained to general writing in L1 (Hamp-Lyons, 2006) or in L2 (Gottlieb, 2000), with marginal focus on the L2 learners' genre-based writing performance, their weaknesses and their goals. Therefore, an urge to further research on this topic was strongly felt, particularly in EFL context.

3. This Study

To bridge the gap in the research literature of portfolio assessment, this study investigated the impacts of genre-based portfolio assessment on 46 EFL undergraduate students' engagement in descriptive and narrative writing. The academic goal for choosing descriptive genre of writing was to examine EFL students' ability in describing tables, figures, flowcharts and other course-related writings at university level. Likewise, the academic goal for choosing narrative genre of writing was to observe EFL students' performance on reporting the stepwise experimental procedures or scientific processes in their term projects or scientific reports. Moreover, EFL students' engagement was conceptualized as a 'metaconstruct' or a framework in which the student regular self-assessment, critical thinking, motivation, self-efficacy and enthusiasm were integrated to achieve the learning goals (Fredricks, Blumenfeld, & Paris, 2004).

The research design was a 12-week genre-based PA, in which the distribution of teacher formative assessment was obtained once-a-week on the students' descriptive and narrative written scripts. The student engagement was assessed through the students' submitted reflective journals. To this end, the following research questions were raised:

RQ1: What is the impact of portfolio assessment on EFL learners' descriptive writing?

RQ2: What is the impact of portfolio assessment on EFL learners' narrative writing?

RQ3: What are the EFL learners' perceptions of the portfolio assessment, teacher feedback, and their writing improvement?

4. Method

4.1 Participants

A sample of 46 EFL undergraduate students at different university majors (architecture, economy, mathematics and MBA) took part in this study. They were sophomore students at Karaj Azad University, Iran. A non-random purposive sampling method was adopted in this study (Ames et al., 2019), because selecting a representative group of university students with adequate experience in genre-based writing in

English was the researcher’s plan. The selected participants in this study had already performed on at least 10 genre-based writing tasks, such as writing an invitation card, a shopping list, a letter of application or travel journals as partial requirements in previous English writing courses at university. Their participation was voluntary and no payment was granted to them. Their experience in learning English was between 3 to 6 years (m = 5) and their ages ranged from 20 to 31 (m = 25.5). Their general English proficiency level was measured by administering Oxford Placement Test, and determined as intermediate.

Additionally, four MA graduates of English language teaching (ELT) took part in this research as the EFL teacher (n = 1) and assistant researchers (n = 3) who provided feedback to the students’ descriptive and narrative writing, co-rated them and analyzed the content of their reflective journals. Table 1 summarized the demographic information of the participants.

Table 1
Demographic Information of the Participants

Participant	Gender	Age	University Major	Studying/Teaching English (Year)	OPT Range of Score
EFL Students n = 46	Female, (30) 65%	20–24, (23) 56.5%	Architecture (23) 50%	3–4, (30) 65%	30–33, (31) 67%
		25–31, (20) 43.5%	Economy (11) 24%	5–6, (16) 35%	34–37, (15) 33%
			Mathematics (6) 13%		
			Accounting (6) 13%		
EFL teacher (1)	Female	26	ELT	6	
Assistant Researchers (3)	Female, (2) 66.6%	24–37	ELT	4–7	

4.2 Instruments

4.2.1 Oxford Placement Test (OPT)

OPT has been developed by Oxford University Press and University of Cambridge Local Examinations Syndicate (CLES), which measures the English proficiency levels of the EFL test takers. The ranking measures were provided by Common European Framework of Reference (CEFR) on a six-level scale ranged from A1 (lowest level) to C2 (highest level). In this study, Oxford Placement Test (OPT) (Version 1, 2001) was used as a placement test to determine the EFL students’ knowledge of general English, and to screen and exclude the outliers. The participants were required to answer the 60 items in the test in 45

minutes. Their language proficiency level was measured as intermediate (30–37, B1 in OPT ranking system) (Cronbach's $\alpha = .812$, representing a strong test reliability).

4.2.2 West Virginia Department of Education (WVDE) Descriptive Writing Rubric

The WVDE rubric meets the criteria for assessing descriptive writing in EFL context, by defining reasonable cut-off scores to ensure a reliable impression of student performance in English. WVDE analytic rubric consists of five components of organization, development, sentence structure, word choice and grammar, and mechanics, within the 1–6 band scores, ranging from 1 (Minimal) to 6 (Exemplary) spectrum. 'Organization' entails clear and logical progression of ideas in a descriptive writing. 'Development' is the writer's clear focus on the intended audience through strong use of examples, relevant details, analogies and illustrations. 'Sentence structure' refers to the use of well-constructed sentences of various structures. 'Word choice and grammar' is the writer's choice of vivid words and grammatical phrases. Finally, 'mechanics' refers to the systematic use of punctuation, capitalization and grammar in writing (NBCT Office of Assessment West Virginia Department of Education, 2015). In this study, WVDE descriptive writing rubric was selected for both teacher feedback and students' self-assessment of their descriptive writing performance. The reason behind adopting this rubric was its user-friendliness, clarity of rubric indicators, and creditability (Appendix A).

4.2.3 Smarter Balanced Narrative Writing Rubric

In 2014, Smarter Balanced Assessment Consortium (SBAC) and Tulare County Office of Education created a digital library of formative assessment tools to support students and teachers in. Smarter Balanced narrative writing rubric has since been used to assess language learners' narrative writing. This analytic rubric consists of five categories of narrative focus, organization, elaboration of narrative, language and vocabulary and conventions. 'Narrative focus' refers to the writer's effective establishment of a setting, narrator and/or characters. 'Organization' entails creating an effective plot which demands unity and completeness. 'Elaboration of narrative' refers to the narrator's thorough and effective elaboration of a narration by using details and dialogues. 'Language and vocabulary' refer to the writer's deliberate choice of words and structures that express personal experience or events. Finally, 'convention' indicates the effective and consistent use of punctuation, capitalization, and spelling in narration (Smarter Balanced Assessment Consortium, 2012). The Smarter Balanced narrative rubric has a 1–5 band scores, ranging from 0 (No evidence of the ability to write a narrative) to 5 (Meeting all the criteria of writing a real or imagined narrative). In this study, teacher feedback and students' self-assessment of narrative writing were conducted with reference to Smarter Balanced narrative writing rubric. The logic behind selecting this rubric was its clear-cut band scores and user-friendliness which made it more accessible to students' self-assessment (Appendix B).

4.2.4 Descriptive and Narrative Writing Elicitation Tasks

For 12 weeks, the participants were required to write and revise their descriptive and narrative drafts through their self-assessment, and regular teacher feedback. To select and incorporate 12 most favorable

topics into descriptive and narrative writing elicitation tasks, a topic familiarity checklist was prepared by the researcher and distributed among the participants. To maintain the variety, the researcher split the descriptive and narrative writing tasks and assigned them into odd and even weeks. For summative assessment purposes, the initial descriptive and narrative tasks in Weeks 1 and 2 were labeled as the descriptive and narrative pretests, and the final descriptive and narrative tasks in Weeks 11 and 12 were named as the descriptive and narrative post-tests. The topics and arrangement of the tasks were summarized in Table 2.

Table 2
Topics and Arrangement of the Descriptive and Narrative Writing Elicitation Tasks

Session	Genre	Topic
1	Descriptive	What are your reading habits? (<i>Descriptive Pretest</i>)
2	Narrative	What was your best vacation? (<i>Narrative Pretest</i>)
3	Descriptive	How do you play with your smartphone?
4	Narrative	What was the last time you went to a museum?
5	Descriptive	What type of games do you like most?
6	Narrative	What do you remember from your childhood?
7	Descriptive	How do you like your university?
8	Narrative	What do you remember from the last wedding reception you were invited?
9	Descriptive	How do you look for a mobile application?
10	Narrative	What does your life look like without your family?
11	Descriptive	How do you clean your room? (<i>Descriptive Post-test</i>)
12	Narrative	What do you remember from the worst day of your life? (<i>Narrative Post-test</i>)

4.2.5 Student Reflective Journals

Reflective journals were written by the students after receiving the teacher feedback on their every written script and revising it accordingly. They were required to respond to four reflective prompts in their reflective journal about the writing task of the week. The prompts were prepared by the researcher and distributed in hard copies in order to collect data about how the participants engage in their (1) experience with writing portfolio system, (2) understanding of received teacher feedback, (3) ability to use teacher feedback in their revised drafts, and (4) perception of writing improvement. The students were free to write in English or Persian (students' L1). As a result, in the submitted reflective journals, around 700 words were collaboratively translated into English by the assistant researchers. The word limit in reflective journals was 500. Therefore, a corpus of around 276,000 words (46*500*12) was submitted to document analysis by the assistant researchers (Cohen's kappa (κ) = .830, interpreted as strong inter-rater

reliability (Cohen, 1960)). From the written responses to every prompt, frequent themes were extracted and counted every time a similar word or concept was encountered during the content analysis.

5. Data Collection Procedure

The logistics of writing PA comprises four steps of collection, selection, reflection, and teacher delayed evaluation (Lam, 2013). However, the researchers are allowed to modify this framework to make it compatible to the purpose of the research or to cope with other limiting contextual factors (Hamp-Lyons & Condon, 2000). In a typical portfolio, 'collection' is the gradual compilation of students' multiple written drafts. 'Selection' is the student self-collection of best pieces of work for teacher's final grading. Usually, in terms of reflective essays, 'reflection' is the student self-assessment and self-reflection of their own personal and learning experience. 'Delayed evaluation' is assigning grades on the final written drafts by the teacher. In this study, 'selection' was deliberately omitted to collect as much data on reflective journals as possible.

Two days before the study began, an OPT was administered as the placement test in order to normalize the selected participants for their English proficiency level. A day before the research commenced, the researcher provided all the participants (including the EFL teacher and assistant researchers) with a 8-hour tutorial on (1) the frameworks of descriptive and narrative writing by presenting two anchor essays, (2) the two selected rubrics for descriptive and narrative writing assessment (i.e., WVDE descriptive writing rubric and Smarter Balanced narrative writing rubric), and (3) the process of writing reflective journals by responding to the four prompts set for all the writing tasks in the portfolio system. The EFL teacher, assistant researchers and students and were presented with a brief discussion to the nature of teacher feedback in terms of comments, evaluation or suggestions they could give/receive without assigning grades to the writings.

The 12-week writing course was divided into six odd sessions devoted to descriptive writing tasks and six even sessions to narrative writing tasks, every second week. The participants were required to write a 300-word essay on the assigned topic of the week, followed by their self-assessment before they submitted their written draft to the teacher. They were allowed to consult the selected rubrics during their writing and self-assessment. Drafting and self-assessment lasted for 60 minutes. At the end of every session, the teacher collected the papers and provided her handwritten comments, corrections or suggestions with reference to the selected rubrics, and in collaboration with assistant researchers. Every comment, correction or suggestion made on the student written drafts was considered as one 'feedback point' and the total feedback points for every draft of writing were calculated. The commented papers were returned to the students in the following week. The students were required to revise their first draft according to the received feedback, and to write a 500-word reflective journal on their personal and learning experience, in terms of responses to the prompts. The revised drafts and reflective journals were stored by the participants for their portfolio compilation and teacher delayed evaluation. By the end of the course, the teacher evaluated the portfolios in holistic approach by assigning them the letters A, B or C, based on the overall quality of the revised final drafts and completeness of the submitted portfolios.

To summatively assess the student writing performance, the assistant researchers scored the descriptive and narrative pre- and post-tests by counting the feedback points, with reference to the writing rubrics. The weekly teacher feedback, summative assessment of the student genre-based writing, and content analysis of the reflective journals were carried out collaboratively by the EFL teacher and assistant researchers. The whole process was supervised by the researcher in the study. In cases of rating or coding disagreement, ongoing negotiation was carried out until agreement was reached upon every occasion. (See Appendix C and D for selected samples of descriptive and narrative writing, along with teacher feedback). The inter-rater reliability indices were calculated for a variety of ratings in this study (Cronbach's $\alpha = 0.800$, reliability index for descriptive writing tasks; Cronbach's $\alpha = 0.981$, reliability index for narrative writing tasks; Cronbach's $\alpha = 0.881$, reliability index for descriptive writing pre and post-test scores; Cronbach's $\alpha = 0.931$, reliability index for narrative writing pre and post-test scores). All values of Cronbach's alpha represented strong agreement and statistical significant ($p < .05$).

The counted feedback points on students writing were keyed into Statistical Package for Social Sciences (SPSS) Version 25. The researcher conducted descriptive statistics to explore the normality of the data, two separate paired-samples t-tests to examine the student performance on pre- and post-test of descriptive and narrative writing, and repeated measures ANOVA to identify significant trends in teacher feedback points on the students' descriptive and narrative writing in the 12-week portfolio assessment. It should be noted that, the observed *decrease* in teacher feedback points was interpreted as the student progress in their writing. The significance level was set at 0.05 in this study. To analyze and interpret the collected qualitative data, the EFL teacher and assistant researchers carried out the inductive content analysis of the compiled reflective journals over the course of 12 weeks. They schematized and reported the extracted themes out of the responses to the prompts ($n = 4$) set for every descriptive and narrative writing task.

6. Results

6.1 Impact of PA on Students Descriptive Writing

Before running statistical tests, descriptive statistics and the assumption of normality were examined for the feedback points on six tasks of descriptive writing (Table 3).

Table 3
Descriptive Statistics and Test of Normality; Descriptive Writings

Week	Min	Max	Mean	Std. Deviation	Skewness		Kurtosis	
					Statistic	Std. Error	Statistic	Std. Error
1(Pretest)	18	92	19.19	8.167	1.386	.564	1.366	1.091
3	20	88	20.31	8.522	.870	.564	1.649	1.091
5	8	88	17.69	10.005	1.027	.564	1.048	1.091
7	30	66	21.81	6.145	.858	.564	-.231	1.091
9	24	54	15.88	4.225	1.197	.564	1.629	1.091
11(Post-test)	20	66	17.44	6.271	.990	.564	.842	1.091

Displayed in Table 3, the mean of feedback points decreased from Task 1 ($m = 19.19$) to Task 11 ($m = 17.44$), which was interpreted as slight improvement in student performance on descriptive writing tasks. The dispersion of the feedback points was also narrowed from Task 1 ($SD = 8.167$) to Task 11 ($SD = 6.271$), which suggested an increasing uniformity in the students descriptive writing performance along the course. The assumption of the normality of the data was not violated, since the measures of skewness and kurtosis fell within the range of ± 2.00 (George & Mallery, 2010). To statistically examine the differences between the students' performance on descriptive pretest in Week 1 and post-test in Week 11, the researcher calculated a paired-samples t-test (Table 4).

Table 4
Paired-Samples t-test; Pre-test and Post-test of Descriptive Writing

Paired Differences					<i>t</i>	<i>df</i>	Sig. (2-tailed)
Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
			Lower	Upper			
1.750	4.782	1.195	-.798	4.298	1.464	45	.164

As seen in Table 4, the results of the paired-samples t-test were insignificant ($t(45) = 1.464$, $p = .164 > .05$, 95% CI [-.798, 4.298], $r = 0.224$, representing a small effect size (Lenhard & Lenhard, 2016)). The findings could be interpreted as the low effect of PA on students descriptive writing improvement. In order to graphically display the weekly trend of the feedback points on student descriptive writings, Fig. 1 was drawn with reference to the five components in the WVDE descriptive writing rubric.

As Fig. 1 shows, the jam of feedback points were split up. At the top, two components of 'mechanics' and 'sentence structures' had a highly fluctuating trend, and at the bottom, the other three components of

'word choice and grammar', 'organization', and 'development' showed a similar unsystematic trend. On the other hand, while the pattern of feedback points on the three latter components had a gradual downfall, the two latter components had a rise-up, by the end of the six-week period. As mentioned before, the decrease in the feedback points was interpreted as the sign of student writing improvement.

Since it was not clear whether the fluctuations of the feedback points on the five components of descriptive writing were meaningful over the six-week period, a set of one-way repeated measures (RM) ANOVA was conducted (Table 5). Before running RM ANOVA, the researcher calculated Mauchly's Test of Sphericity for the data which indicated that the assumption of sphericity was not violated, $\chi^2(44) = .142$, $p = .627 > .05$.

Table 5
Repeated Measures ANOVA for Components of Descriptive Writing: Six Week Period

	Sum of Squares	df	Mean Square	F	Sig.	Partial η^2
<i>Organization</i>						
Between Group	508.760	1	508.760	195.312	.000	.929
Within Group	12.677	5	2.535	2.486	.039	.142
<i>Development</i>						
Between Group	352.667	1	352.667	88.659	.000	.855
Within Group	126.708	5	25.342	10.741	.000	.417
<i>Sentence Structure</i>						
Between Group	748.167	1	748.167	331.700	.000	.957
Within Group	33.833	5	6.767	3.054	.015	.169
<i>Word choice & Grammar</i>						
Between Group	4056.000	1	4056.000	74.075	.000	.832
Within Group	48.875	5	9.775	2.015	.036	.118
<i>Mechanics</i>						
Between Group	2948.167	1	2948.167	52.037	.000	.776
Within Group	91.208	5	18.242	1.893	.106	.112

As Table 5 illustrates, the teacher feedback points on all five components of descriptive writing showed significant differences, including the weekly mean scores of 'organization' (F (1, 5) = 195.312, $p = .000 < .05$, $\eta^2 = .929$), 'development' (F (1, 5) = 88.659, $p = .000 < .05$, $\eta^2 = .855$), 'sentence structure' (F (1, 5) = 331.700, $p = .000 < .05$, $\eta^2 = .957$), 'word choice & grammar' (F (1, 5) = 74.075, $p = .000 < .05$, $\eta^2 = .832$),

and 'mechanics' ($F(1, 5), 52.037, p = .000 < .05, \eta^2 = .776$). All the measures of partial η^2 were interpreted as strong effect sizes (Lenhard, & Lenhard, 2016).

6.2 Impact of PA on the Students Narrative Writing

Once again, descriptive statistics and the assumption of normality for the feedback points on six tasks of narrative writings were tested (Table 6).

Table 6
Descriptive Statistics and Test of Normality; Narrative Writings

Week	Min	Max	Mean	Std. Deviation		Skewness		Kurtosis	
				Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error
2 (Pretest)	42	94	28.25	9.581	1.176	.564	-.108	1.091	
4	34	80	25.88	6.313	.563	.564	.207	1.091	
6	36	140	26.38	12.675	1.009	.564	0.376	1.091	
8	20	156	20.94	15.889	1.861	.564	1.821	1.091	
10	20	72	18.69	7.922	.863	.564	.234	1.091	
12 (Post-test)	24	64	18.06	5.859	.950	.564	.387	1.091	

As seen in Table 6, the mean of feedback points largely decreased in number, from Task 2 ($m = 28.25$) to Task 12 ($m = 18.06$), which was interpreted as notable progress in students' performance on narrative writing tasks. The dispersion of the feedback points was also largely shrunk from Task 2 ($SD = 9.581$) to Task 12 ($SD = 5.859$), to show an increasing homogeneity in the students' narrative writing. In terms of normal distribution of the data, the measures of skewness and kurtosis were within the range of ± 2.00 (George & Mallery, 2010), which maintained the normality of the data. To examine the progress in the students narrative writing, the researcher conducted a paired-samples t-test between the students' performance on narrative pretest in Week 2 and post-test in Week 12.

Table 7
Paired-Samples t-test; Pre-test and Post-test of Narrative Writing

Paired Differences					<i>t</i>	<i>df</i>	Sig. (2-tailed)
Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
			Lower	Upper			
10.188	11.053	2.763	4.298	16.077	3.687	45	.002

In Table 7, the results of paired-samples t-test were significant ($t(45) = 3.687, p = .002 > .05$, 95% CI [4.298, 16.077], $r = 1.281$, representing a large effect size (Lenhard & Lenhard, 2016)). They were interpreted as the positive impact of PA on student narrative writing improvement in six-week period. Figure 2 illustrates the weekly trend of the feedback points on student narrative writing, with reference to five components in the Smarter Balanced narrative writing rubric.

In Fig. 2, the feedback points polarized the component of 'narrative focus' at the top, while other four components of 'organization', 'elaboration of narrative', 'language and vocabulary', and 'convention' received less feedback points and were marginalized at the bottom. The weekly trend of 'narrative focus' showed a positive rise, while the rest of four components had a less fluctuating trend and minimized by the end of six-week period, which indicated improvement of those qualities in student narrative writing. Yet, the paired-samples t-test results were not able to account for the significance of fluctuations in five components of narrative writing. Therefore, another set of one-way repeated measures ANOVA was run (Table 8). The measure of Mauchly's Test of Sphericity for the data indicated that the assumption of sphericity was not violated, $\chi^2(44) = .088, p = .505 > .05$.

Table 8
Repeated Measures ANOVA for Components of Narrative Writing: Six Week Period

	Sum of Squares	df	Mean Square	F	Sig.	Partial η^2
<i>Narrative focus</i>						
Between Group	1488.375	1	1488.375	64.908	.000	.812
Within Group	372.625	5	74.525	8.774	.000	.369
<i>Organization</i>						
Between Group	481.510	1	481.510	774.721	.000	.981
Within Group	22.552	5	4.510	7.105	.000	.321
<i>Elaboration of narrative</i>						
Between Group	10.333	5	2.067	2.598	.032	.848
Within Group	210.042	1	210.042	263.467	.000	.946
<i>Language and vocabulary</i>						
Between Group	870.010	1	870.010	169.139	.000	.919
Within Group	48.552	5	9.710	3.618	.006	.194
<i>Convention</i>						
Between Group	15887.760	1	15887.760	90.326	.000	.858
Within Group	145.177	5	29.035	1.253	.013	.077

In Table 8, significant differences can be observed in the teacher feedback points on all five components of narrative writing, including the weekly mean scores of 'narrative focus' ($F(1, 5) = 64.908, p = .000 < .05, \eta^2 = .812$), 'organization' ($F(1, 5) = 774.721, p = .000 < .05, \eta^2 = .981$), 'elaboration of narrative' ($F(1, 5) = 2.598, p = .032 < .05, \eta^2 = .848$), 'language and vocabulary' ($F(1, 5) = 169.139, p = .000 < .05, \eta^2 = .919$), and 'convention' ($F(1, 5) = 90.326, p = .000 < .05, \eta^2 = .858$). All the measures of partial η^2 were interpreted as strong effect sizes (Lenhard, & Lenhard, 2016).

6.3 Analysis of the Student Reflective Journals

An inductive content analysis was carried out with the 46 EFL students' self-reported reflective journals over the 12 week of portfolio assessment. As Table 9 summarizes, the inductive analysis of the responses to the given prompts resulted in eleven themes. The themes were extracted and categorized according to the responded prompts. The majority of responses to the first prompt, which asked for the students' reflection on their learning and personal experience with PA, agreed upon the successful and positive impact of the PA (62.5%). 33.68% of the responses pointed to the novelty of their experience with PA, and only 3.81% of them expressed their frustration with working in portfolio system.

The encoded themes for the second prompt summarized a large proportion of the student critical engagement in the teacher feedback. 54.12% of the responses expressed an unexpected mismatch between the student self-assessment and the received teacher feedback, both in number, in type and in feedback focus. More specifically, the students observed a sensitivity or bias in teacher assessment towards certain form-focused components in their writing such as mechanics or choice of words, at the expense of feedback to macro-components of organization or development of ideas (45.87%). Students believed that such tendency in teacher feedback narrowed down their focus to prioritize improving certain types of committed errors.

To the third prompt which asked for the student reflections on the applicability of teacher feedback in their revision process, almost half of the responses agreed on its usefulness (50.13%). The students also were satisfied with their experience of learning discourse features such as cohesive devices, generating 'good ideas' in writing, and their gradual alignment to genre conventions. Yet a large number of responses pointed to the teacher comments as incomprehensible, difficult to apply (26.38%), or inadequate (23.48%). The last prompt required the students' self-evaluation of their writing progress, to which the majority of the students positively responded (74.16%). 21.03% of the responses pointed to partial improvement of their writing, and only 4.79% of them found portfolio assessment as ineffective to their writing betterment.

Table 9
Extracted Themes; Distribution in Student Reflective Journals

<i>Prompt 1: What do you think about keeping a writing portfolio in general?</i>		
Exemplar response	Theme	F (Percentage)
- I like it that I have a chance to correct my mistakes.	Effectiveness	180 (62.5%)
- Interesting! I haven't done it before.	Novelty	97 (33.68%)
- I don't like it much because I will revise my writing again and again!	Uselessness	11 (3.81%)
<i>Prompt 2: What do you think about the (quality and/or quantity) of the teacher feedback on your draft?</i>		
- Usually, I get more feedback than I expect.	Mismatch between self-assessment and teacher feedback	118 (54.12%)
- I think she [the teacher] is more sensitive to my good words than my good ideas!	Feedback bias	100 (45.87%)
<i>Prompt 3: Could you use the teacher feedback when you revised your draft?</i>		
- Her comments are very useful.	Usefulness	190 (50.13%)
- I don't understand some of her comments.	Inapplicability	100 (26.38%)
- I need more help on this assignment.	Inadequacy	89 (23.48%)
<i>Prompt 4: What do you think about your writing progress this week?</i>		
- I have learned a lot this week.	Positive effect	201 (74.16%)
- I can see some improvement.	Moderate effect	57 (21.03%)
- I don't see any change yet.	No effect	13 (4.79%)

7. Discussion

In line with the purpose of the study, three research questions were raised and explored. In research question 1, the researcher's purpose was to gain insight into the impact of portfolio assessment on student descriptive writing. The analysis of teacher summative assessment indicated no evidence for positive impact of PA on students' descriptive performance. However, further statistical results of one-way repeated measures ANOVA indicated the meaningful impact of PA on the student descriptive writing process, despite no observable improvement in their written product on the posttest of descriptive writing. The findings implied that despite the insufficient skills and knowledge of genre-based writing, the students could engage in self-assessment, apply the received teacher feedback, and critically evaluate the

quality of their writing from a descriptive-genre perspective. In other words, in the process of genre-based portfolio assessment, the students had a chance to receive feedback on both microscopic (such as 'mechanics' of writing or, 'sentence structure') and macroscopic (such as 'organization' or 'development' of ideas) aspects of descriptive writing (Borg, 2003).

In a case study of the challenges an Indonesia EFL teacher might face in portfolio assessment, Halim and Lestari (2019) also reported improvement in students' descriptive writing despite their low rate of engagement and the teacher's difficulties in supervising the student peer and self-assessment. The findings were in contradiction with Roohani and Taheri (2015) who supported the positive impact of PA on the subskills in student descriptive and expository writing achievement. However, their reported impact on the students' choice of words and conventions of writing was interpreted as weak and temporary.

In response to research question 2 which explored the impact of portfolio assessment on student narrative writing, the analysis of data from teacher formative and summative assessment indicated the positive double impacts of PA on improving the student narrative writing process, and on final products in terms of their writing performance on the posttest of narrative writing. In other words, the students' writing progress and final achievement suggested their constant reference to the multiple components in the selected writing rubric (i.e., Smarter Balanced narrative writing rubric), close observation of the received teacher commentaries, and successful and systematic application of them to their revised scripts.

The research literature on portfolio assessment and narrative writing dated back to the 1990s. In a case study with 22 EFL students, Shober (1996) conducted 12-week portfolio assessment and reported contrary results that only 68 percent of the students demonstrated improvement in narrative writing. Twenty-seven percent of the students' scores remained unchanged, and a single student had 5 percent decrease in her final score. Shober (1996) concluded that portfolio assessment was deficient and ineffective as an evaluation tool. In another case study, Gearhart et al. (1992) adopted a methodological approach to portfolio assessment of 35 English-speaking elementary students' narrative writing. They reported critical issues regarding efficiency of portfolio assessment as an approach to evaluate students' narrative writing, such as controversial 'scorability' of the portfolio, disagreements over the 'domains for portfolio assessment', and its 'utility for large-scale assessment'.

Research question 3 explored the degree of student engagement in portfolio assessment of their descriptive and narrative writing performance. Regarding their general perception of writing portfolio assessment, the majority of the students agreed upon the merits and novelty of their experience. In the same vein, findings by Herberg (2005), Pollari (2000), and Song and August (2002) showed the positive attitudes of students to PA as a learning tool for EFL writing. The findings in student reflective journals were also in line with several studies that reported the impacts of writing PA on the students' confidence, motivation and positive learning attitude (Afrianto, 2017; Ainley, 2012; Black & William, 2018; Charanjit & Samad, 2013; Lam, 2019; Steen-Utheima & Hopfenbeck, 2019; Zhao, 2010). However, in a number of

qualitative case studies, portfolios were underrated by some students as inefficient, confusing and challenging (Boyden-Knudsen, 2001; Hirvela & Sweetland, 2005, Price et al., 2010; Zhang & Hyland, 2018).

Regarding the student engagement in the effectiveness of teacher feedback, the students reported an incompatibility between the teacher feedback and their self-assessment both in number (e.g., teacher feedback was outnumbered) and in nature (e.g., teacher feedback was more detailed, confusing and repetitive). The findings were in line with several studies that reported the student failure in making sense of teacher feedback and, their subsequent low attention and required action (Carless, 2011; Pierce et al., 2010). The students also brought up the issue of teacher feedback *sensitivity*. They reported teacher sensitivity towards certain writing features, such as 'punctuation', 'description of setting in narrative', 'choice of words', and systematic leniency towards others, such as 'development of supporting ideas' or 'bringing details or examples'. The self-reports in student reflective journal were closely in line with the concept of 'didactical contract' (Brousseau, 1984). Regarding the assessment practice as a social process, the students and the teacher seem like actors interacting inside a network of mutual expectations where the students tend to 'legitimize' learning and laser-focus those areas in the new language which are bolded in teacher feedback. For example, in a classroom where the teacher's comments often center around the production of correct sentence structures or cosmetic features of language, students may interpret future feedback on the development of coherent ideas as 'illegitimate' or 'unfair'.

About half of the student perceptions on their ability to use teacher feedback in their revised drafts was positive and satisfactory. The other half pointed to the inapplicability and inadequacy of the teacher feedback. Several other studies on PA reported either the students misinterpreted teacher comments when revising their written drafts (Belgrad, 2013; Clarke & Boud, 2016; Price et al., 2010; Swaffield, 2011), or failed to revise them after several re-drafting (Ferris, 1997). Finally, the majority of the participants in this study reported their writing improvement while participating in the writing portfolio system. The findings were supported in a number of studies in which the L2 students perceived noticeable writing achievements in a portfolio system (Aydin, 2010; Song & August, 2002), and contradicted in a few others (Hamp-Lyons & Condon, 2000; Lam & Lee, 2010).

8. Conclusion

The current study explored how genre-based portfolio assessment could impact the EFL undergraduate students descriptive and narrative writing performance and how they could engage, perceive, and act upon teacher feedback in the 12-week portfolio assessment platform. It was a case study on 46 EFL learners with some limitations in results and implications for future researchers and L2/EFL teachers. In this study, the participants were selected with non-random purposive sampling to participate in a case study. Consequently, the generalizability of the findings in this research will be limited; yet the insight of how genre-based PA might affect the EFL learners writing at university level can provoke further research in educational settings of colleges and universities. Secondly, the collected data were limited to teacher formative assessment of student writing and the student reflective journals in 12-week portfolio

assessment. The researcher strongly believes that extending the period of data collection could have yielded richer data on the student genre-based writing progress. Thirdly, more critical data could be collected from setting elicitation recall or interviewing the EFL teacher and assistant researchers, who were in charge of giving weekly feedback on the student writings. Fourthly, the students' engagement in PA was restricted to their self-reported perceptions in terms of reflective journals. Further research may add peer assessment or active collaboration of students in drafting and revising their texts, as further important sources of data on language learner engagement. Last but not least, no analytical analysis was conducted on the student revision process and their effective application of received teacher feedback, which can be a demanding topic for future research.

List Of Abbreviations

AfL (Assessment for Learning)

EFL (English as a Foreign Language)

L2 (Second/Foreign Language)

OPT (Oxford Placement Test)

PA (Portfolio Assessment)

ANOVA (Analysis of Variance)

WVDE (West Virginia Department of Education)

Declarations

Availability of data and materials

Please contact the author for data requests.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests

Not applicable.

Author's information

Natasha Pourdana is a PhD in applied linguistics and an assistant professor at KIAU, Iran. Her fields of interest are language assessment, translation quality assessment and computer-assisted language

learning.

References

1. Abbott, P. H. (2002). *The Cambridge introduction to narrative*. Cambridge: Cambridge University Press.
2. Afrianto, A. (2017). Challenges of using portfolio assessment as an alternative assessment method for teaching English in Indonesian schools. *International Journal of Educational Best Practices*, 1(2), 106-114.
3. Ainley, M. (2012). Students' interest and engagement in classroom activities. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 283–302). New York, NY: Springer.
4. Ames, H., Glenton, C. & Lewin, S. (2019). Purposive sampling in a qualitative evidence synthesis: a worked example from a synthesis on parental perceptions of vaccination communication. *BMC Med Res Methodol*, 19(26). <https://doi.org/10.1186/s12874-019-0665->
5. Aydin, S. (2010). EFL writers' perceptions of portfolio keeping. *Assessing Writing*, 15(3), 194–203.
6. Badger, R. & White, G. (2000). A process genre approach to teaching writing, *ELT Journal*, 54(2), 153-160.
7. Belgrad, S. F. (2013). Portfolios and e-portfolios: Student reflection, self-assessment, and global setting in the learning process. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 331–346). Los Angeles: Sage.
8. Benson, P. (2006a). Autonomy and its role in learning. In J. Cummins & C. Davison (eds.), *The international handbook of English language teaching* (vol. 2). Norwell, MA: Springer.
9. Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31.
10. Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551-575.
11. Birjandi, P., & Hadidi Tamjid, N. (2012). The role of self, peer and teacher assessment in promoting Iranian EFL learners' writing performance. *Assessment & Evaluation in Higher Education*, 37(5), 513–533.
12. Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do, *Language Teaching*, 36(2), 81-109.
13. Boyden-Knudsen, T. (2001). The effects of analytic corrections and revisions on college composition students in a portfolio assessment setting. *ERIC*, ED457171
14. Brown, H D, & Abeywickrama, P. (2010). *Language Assessment: Principles and Classroom Practices*. White Plains, NY: Pearson Education.
15. Brousseau, G. (1984). The crucial role of the didactical contract in the analysis and construction of situations in teaching and learning mathematics, In H. G. Steiner (ed.), *Theory of mathematics*

- education*, (pp. 110-119). Bielefeld, IDM.
16. Burner, T. (2014). The potential formative benefits of portfolio assessment in second and foreign language writing contexts: A review of the literature. *Studies in Educational Evaluation*, 43, 139–149.
 17. Caner, A. (2010). Students views on using portfolio assessment an EFL writing courses. *Anadolu University Journal of Social Sciences*, 10(1), 223-236.
 18. Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings*. New York: Routledge.
 19. Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325.
 20. Charanjit, S. & Samad, A. (2013). Developing a portfolio assessment model for the teaching and learning of English in Malaysian I2 classroom. *English Language Teaching*, 8(7), 14-16.
 21. Clarke, J. L., & Boud, D. (2016). Refocusing portfolio assessment: Curating for feedback and portrayal. *Innovations in Education and Teaching International*, doi:10.1080/14703297.2016.1250664
 22. Cohen, J. (1960). "A coefficient of agreement for nominal scales". *Educational and Psychological Measurement*, 20 (1), 37–46.
 23. Condon, W., & Hamp-Lyons, L. (1994). Maintaining a portfolio-based writing assessment: Research that informs program development. In: L. Black, D. A. Daiker, J. Sommers, & G. Stygall (Eds.), *New directions in portfolio assessment: Reflection practice, critical theory, and large-scale scoring* (pp. 277–285). Portsmouth, NH: Boynton/Cook.
 24. Cowie, B. (2005). Pupil commentary on assessment for learning. *Curriculum Journal*, 16(2), 137–151.
 25. Dann, R. (2014). Assessment as learning: Blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice*, 21(2), 149–166.
 26. Darling-Hammond, L., & McCloskey, L. (2008). Assessment for learning around the world: What would it mean to be internationally competitive? *The Phi Delta Kappan*, 90 (4), 263–72.
 27. De Fina, A., & Georgakopoulou, A. (2015). *The handbook of narrative analysis*. UK: Wiley and Blackwell.
 28. Duff, P. A. (2010). Language socialization into academic discourse communities. *Annual review of applied linguistics*, 30, 169.
 29. Earl, L. M. (2013). *Assessment as learning: Using classroom assessment to maximize student learning* (2nd ed.). Thousand Oaks, CA: Corwin.
 30. Ecclestone, K. (2007). Commitment, compliance and comfort zones: the effects of formative assessment on vocational education students' learning careers. *Assessment in Education: Principles. Policy & Practices*, 14(3), 315–333.
 31. Ferris, D. R. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly*, 31(2), 315-339.

32. Finn, J. D., & Zimmer, K. S. (2012). Student Engagement: What Is It? Why Does It Matter? In S. L. Christenson, A. L. Reschly, and C. Wylie, *Handbook of research on student engagement* (pp. 97 -130). New York: Springer.
33. Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74, 59–109.
34. Gearhart, M., Herman, J. L., Baker, E. L., & Whittaker, A. K. (1992). *Writing portfolios at the elementary level: A study of methods for writing assessment* (CSE Report 337). Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
35. George, D., & Mallery, M. (2010). *SPSS for Windows step by step: A simple guide and reference*, 17.0 update (10th ed.) Boston: Pearson.
36. Gottlieb, M. (2000). Portfolio practices in elementary and secondary schools. In G. Ekbatani & H. Pierson (Eds.), *Learner-directed assessment in ESL* (pp. 89-104). Mahwah, NJ: Lawrence Erlbaum.
37. Gregory, K., Cameron, C., & Davies, A. (2001). *Knowing what counts: Conferencing and reporting*. Merville, BC: Connections Publishing.
38. Halim, I. A., & Lestari, Z. W. (2019). the use of portfolio assessment of writing skill in descriptive text. *Journal and Applied Linguistics and Literacy*, 3(2), 75-85.
39. Hamp-Lyons, L. (2006). Feedback in portfolio-based writing courses. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing contexts and issues* (pp. 140–161). London: Cambridge University Press.
40. Hamp-Lyons, L. (2007). The impact of testing practices on teaching: Ideologies and alternatives. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching* (pp. 487–504). Norwell, MA: Springer.
41. Hamp-Lyons, L. (2016). Purposes of assessment. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 13–27). Germany: De Gruyter.
42. Hamp-Lyons, L., & Condon, W. (2000). *Assessing the portfolio: Issues for research, theory and practice*. Cresskill, NJ: Hampton Press.
43. Hargreaves, E. (2005). Assessment for learning? Thinking outside the (black) box. *Cambridge Journal of Education*, 35(2), 213–224.
44. Harris, L., & Brown, G. (2009). The complexity of teachers' conceptions of assessment: tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice*, 16(3), 365–381.
45. Herberg, E. (2005). Can a metamorphosis be quantified? Reflecting on portfolio assessment, *Composition Studies*, 33 (2), 69–87.
46. Hinkel, E. (2002) *Second language writers' text: Linguistic and rhetorical features*. Lawrence Erlbaum Associates.

47. Hirvela, A., & Sweetland, Y. L. (2005). Two case studies of L2 writers' experiences across learning directed portfolio contexts. *Assessing Writing*, 10 (3), 192–213.
48. Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing*, 12 (1), 17–29.
49. Hyland, K. (2004). *Genre and second language writing*. Ann Arbor, MI: The University of Michigan Press.
50. Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing*, 16 (3), 148–164.
51. Hyland, K., & Hyland, F. (Eds.). (2006). *Feedback in second language writing: Contexts and issues*. Cambridge, UK: Cambridge University Press.
52. Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy and Practice*, 16(3), 263–268.
53. Labov, W. (2001). *Uncovering the event structure of narrative*. Georgetown University Round Table.
54. Lam, R. (2013). Two portfolio systems: EFL students' perceptions of writing ability, text improvement, and feedback. *Assessing Writing*, 18(2), 132–153.
55. Lam, R. (2014). Promoting self-regulated learning through portfolio assessment: Testimony and recommendations. *Assessment & Evaluation in Higher Education*, 39(6), 699–714.
56. Lam, R. (2017). Taking stock of portfolio assessment scholarship: From research to practice. *Assessing Writing*, 31, 84–97.
57. Lam, R. (2018a). Promoting self-reflection in writing: A showcase portfolio approach. In A. Burns & J. Siegel (Eds.), *International perspectives on teaching skills in ELT* (pp. 219–231). London: Palgrave MacMillan.
58. Lam, R. (2018b). Teacher learning of portfolio assessment practices: Testimonies of two writing teachers. In M. F. Hill & H. Jiang (Eds.), *Teacher learning from classroom assessment: Perspectives from Asia Pacific* (pp. 99–118). New York: Springer.
59. Lam, R. (2019). Writing portfolio assessment in practice: individual, institutional, and systemic issues, *Pedagogies: An International Journal*, DOI: 10.1080/1554480X.2019.1696197
60. Lam, R., & Lee, I. (2010). Balancing the dual functions of portfolio assessment. *ELT Journal*, 64(1), 54–64.
61. Lee, I. (2016). Teacher education on feedback in EFL writing: Issues, challenges, and future directions. *TESOL Quarterly*, 50(2), 518–527.
62. Lee, I., & Coniam, D. (2013). Introducing assessment for learning for EFL writing in an assessment of learning examination-driven system in Hong Kong. *Journal of Second Language Writing*, 22 (1): 34–50.
63. Lenhard, W., & Lenhard, A. (2016). Calculation of Effect Sizes. Retrieved from: https://www.psychometrica.de/effect_size.html. Dettelbach (Germany): Psychometrica. DOI: 10.13140/RG.2.2.17823.92329

64. Li, Q. (2010). The impact of portfolio-based writing assessment on EFL writing development of Chinese learners. *Chinese Journal of Applied Linguistics*, 33 (2), 103–116.
65. Lou-Conlin, M. (1998). *Patterns: A short prose reader*. Boston. New York.
66. NBCT Office of Assessment West Virginia Department of Education (2015). Artificial intelligence scoring of student essays: West Virginia's experience.
67. O'Shea, A. (2019). Teachers' Conceptions of Assessment for Learning: What are the Implications for Children? *Polish Journal of Educational Studies*, II (LXXII) DOI: 10.2478/poljes-2019-0005.
68. Mohamadi, Z. (2018). Comparative effect of online summative and formative assessment on EFL student writing ability. *Studies in Educational Evaluation*, 59, 29-40.
69. Panadero, E. & Romeo, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education Principles Policy and Practice*, 21(2),133-148.
70. Pollari, P. (2000). *This Is My Portfolio: Portfolios in Upper Secondary School English Studies*. Institute for Educational Research, Customer Services, University of Jyvaskyla.
71. Price, M., Handley, K., Millar, J. & O'Donovan, B. (2010). Feedback: all that effort, but what is the effect? *Assessment & Evaluation in Higher Education*, 35(3), 277–289
72. Ramaprasad, A. (1983). On definition of feedback. *Behavioral Science*, 28(1), 4-13.
73. Romova, Z., & Andrew, M. (2011). Teaching and assessing academic writing via the portfolio: Benefits for learners of English as an additional language, *Assessing Writing*, 16, 111-122.
74. Roohani, A. & Taheri, F. (2015). The effect of portfolio assessment on EFL learners' expository writing ability, *Iranian Journal of Language Testing*, 5(1), 45-59.
75. Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35(5), 535–550.
76. Shober, L. S. (1996). A portfolio assessment approach to narrative writing with the cooperation of a fourth grade, *ERIC*, ED395318. SmarterBalanced.org (2020). The Smarter Balanced assessment system propels instruction forward.
77. Song, B., & August, B. (2002). Using portfolios to assess the writing of ESL students: a powerful alternative? *Journal of Second Language Writing*, 11 (1), 49–72.
78. Steen-Utheima, A. & Hopfenbeck, T. N. (2018). To do or not to do with feedback: A study of undergraduate students' engagement and use of feedback within a portfolio assessment design, *Assessment & Evaluation in Higher Education*, <https://doi.org/10.1080/02602938.2018.1476669>
79. Swaffield, S. (2011). Getting to the heart of authentic assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 18(4), 433–449.
80. William, D. (2006) Formative Assessment: Getting the Focus Right. *Educational Assessment*, 11(3–4), 283–289.
81. Willis, J. (2011). Affiliation, autonomy and assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 18(4), 399–415.

82. Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162.
83. Zhang, Z. V. & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing, *Assessing Writing*, 36, 90-102.
84. Zhao, H. (2010). Investigating learners' use and understanding of peer and teacher feedback of writing: A comparative study in a Chinese writing classroom. *Assessing Writing*, 15 (1), 3–17.

Figures

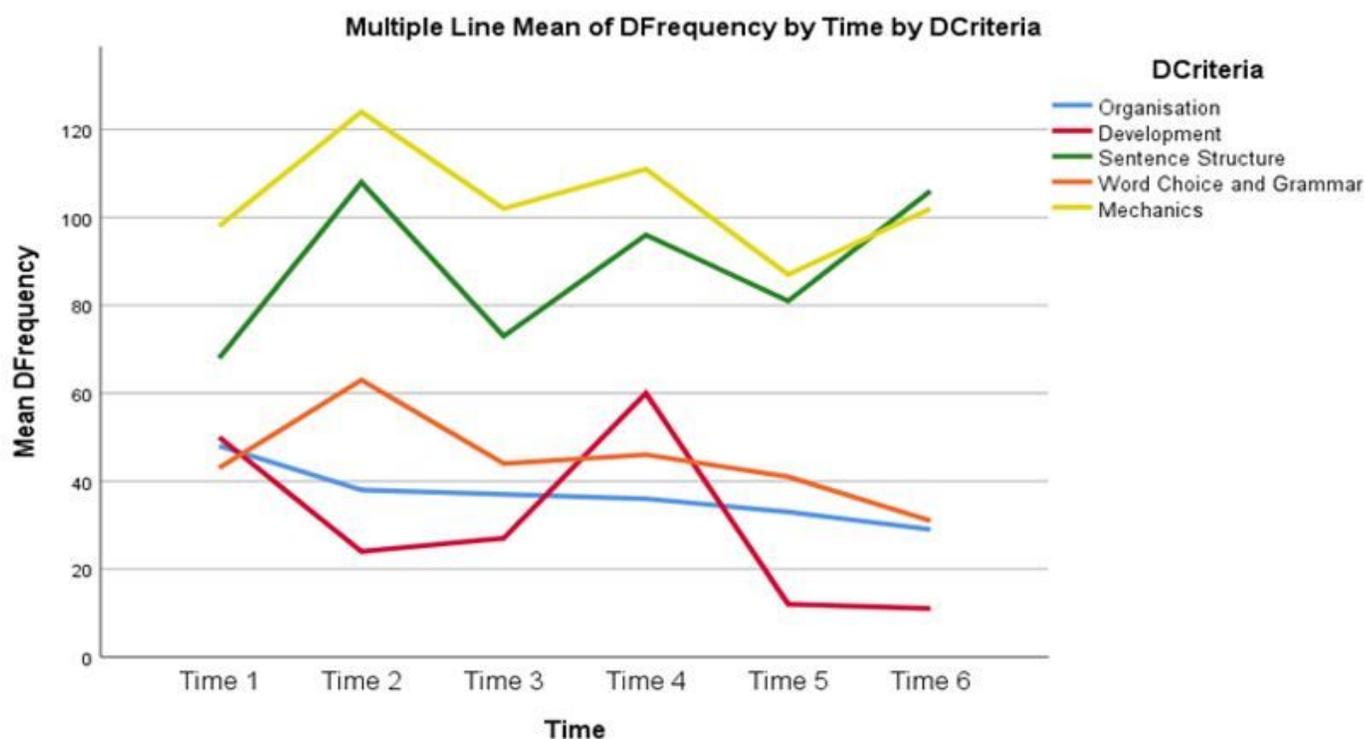


Figure 1

Teacher Formative Assessment of Student Descriptive Writing Over Six-week Period

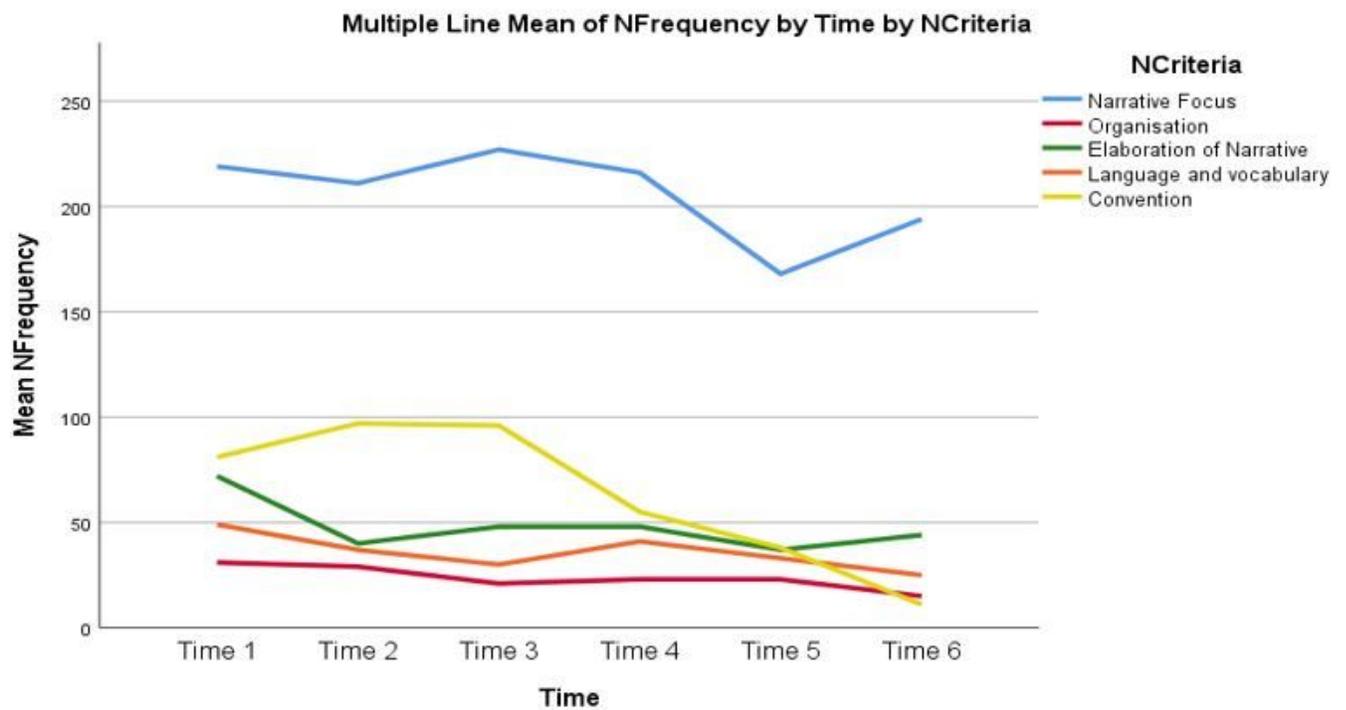


Figure 2

Teacher Formative Assessment of Student Narrative Writing Over Six-week Period

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Appendixes.docx](#)