

What If It's All an Illusion? To What Extent Can We Rely on Self-Reported Data in Open, Online, and Distance Education Systems?

Yavuz Akbulut, Abdullah Saykılı, Aylin Öztürk and Aras Bozkurt

Volume 24, Number 3, August 2023

URI: <https://id.erudit.org/iderudit/1106769ar>
DOI: <https://doi.org/10.19173/irrodl.v24i3.7321>

[See table of contents](#)

Publisher(s)

Athabasca University Press (AU Press)

ISSN

1492-3831 (digital)

[Explore this journal](#)

Cite this article

Akbulut, Y., Saykılı, A., Öztürk, A. & Bozkurt, A. (2023). What If It's All an Illusion? To What Extent Can We Rely on Self-Reported Data in Open, Online, and Distance Education Systems? *International Review of Research in Open and Distributed Learning*, 24(3), 1–17. <https://doi.org/10.19173/irrodl.v24i3.7321>

Article abstract

Online surveys are widely used in social science research as well as in empirical studies of open, online, and distance education. However, students' responses are likely to be at odds with their actual behavior. In this context, we examined the discrepancies between self-reported use and actual use (i.e., learning analytics data) among 20,646 students in an open, online, and distance education system. The ratio of consistent responses to each of the 11 questions ranged from 43% to 70%, and the actual access to learning resources was significantly lower than self-reported use. In other words, students over-reported their use of learning resources. Females were more likely to be consistent in their responses. Frequency of visits to the open, online, and distance education system, grade point average, self-reported satisfaction, and age were positively correlated with consistency; students' current semester was negatively correlated with consistency. Although consistency was not maintained between actual use and self-reported use, consistency was maintained between some of the self-report questionnaires (i.e., use vs. satisfaction). The findings suggested that system and performance data should be considered in addition to self-reported data in order to draw more robust conclusions about the accountability of open, online, and distance education systems.



August – 2023

What If It's All an Illusion? To What Extent Can We Rely on Self-Reported Data in Open, Online, and Distance Education Systems?

Yavuz Akbulut¹, Abdullah Saykılı², Aylin Öztürk², and Aras Bozkurt²

¹Anadolu University, Department of Educational Sciences, Eskisehir, Türkiye; ²Anadolu University, Department of Distance Education, Eskisehir, Türkiye

Abstract

Online surveys are widely used in social science research as well as in empirical studies of open, online, and distance education. However, students' responses are likely to be at odds with their actual behavior. In this context, we examined the discrepancies between self-reported use and actual use (i.e., learning analytics data) among 20,646 students in an open, online, and distance education system. The ratio of consistent responses to each of the 11 questions ranged from 43% to 70%, and the actual access to learning resources was significantly lower than self-reported use. In other words, students over-reported their use of learning resources. Females were more likely to be consistent in their responses. Frequency of visits to the open, online, and distance education system, grade point average, self-reported satisfaction, and age were positively correlated with consistency; students' current semester was negatively correlated with consistency. Although consistency was not maintained between actual use and self-reported use, consistency was maintained between some of the self-report questionnaires (i.e., use vs. satisfaction). The findings suggested that system and performance data should be considered in addition to self-reported data in order to draw more robust conclusions about the accountability of open, online, and distance education systems.

Keywords: open and distance learning, higher education, self-report, inconsistent responding, learning analytics

Introduction

Surveys are one of the most convenient ways to collect data in social science research. Self-reported learner reflections are considered essential for studying most psychological processes related to human learning, such as motivation, emotions, and metacognition (Pekrun, 2020). They are also used to evaluate the accountability efforts of educational institutions or to inform further policy decisions.

With the increase in Internet access worldwide, conducting online surveys has become one of the most preferred ways to collect data from large populations in a very short period of time. Several factors make online surveys a practical research tool, including the ease of data collection and entry (Evans & Mathur, 2005), the elimination of lack of motivation and low response rates, especially for confidential questions (Gregori & Baltar, 2013), and the ability to expand the geographical scope of the target population and study hard-to-reach individuals (Baltar & Brunet, 2012). Due to the intensive use of technology in the delivery of educational content, open online and distance education processes are often studied through online surveys.

While concerns have often been raised about the decline in the amount of robust educational intervention research (Hsieh et al., 2005; Reeves & Lin, 2020; Ross & Morrison, 2008), systematic reviews of educational technology and distance learning show that researchers often adopt survey design and use questionnaires or scales as a data collection tool and then use the results for descriptive or correlational analyses (Bozkurt et al., 2015; Kara Aydemir & Can, 2019; Küçük et al., 2013; Zhu et al., 2020) with an heavy reliance on positivist paradigm (Kara Aydemir & Can, 2019; Mishra et al., 2009). It is certainly tempting to reach many participants with little effort; however, in some cases, the results of survey designs do not necessarily reflect actual situations. While constructing reliable and valid scales is considered central to robust measurement practices, respondents themselves can be a potential source of measurement error. That is, they may provide inconsistent responses (Castro, 2013), exert insufficient effort in responding (Huang et al., 2015), or alter their responses in socially desirable ways (Chesney & Penny, 2013), all of which result in low-quality data that can bias further hypothesis testing steps (DeSimone & Harms, 2018). In many cases, the proportion of inattentive participants or inconsistent responses within a dataset can be negligible, which does not change the inferences or conclusions of the study (Iaconelli & Wolters, 2020; Schneider et al., 2018). However, there are also cases where pronounced effects on reliability have been found (Chesney & Penny, 2013; Maniaci & Rogge, 2014).

According to Albert Bandura's social cognitive theory, the dynamic and reciprocal interaction of personal factors, environmental factors, and the nature of behavior can predict human learning and development (Bandura, 1977). For example, lack of motivation or effort on the part of participants may lead them to simply provide satisfactory answers rather than answering all survey questions optimally, as this may require considerable cognitive effort (Krosnick, 1991). The primacy and recency of self-report questions (Chen, 2010) or participants' anchoring and adjusting behaviors (Zhao & Linderholm, 2008) may further explain response inconsistencies. More specifically, participants' initial responses to self-report measures may serve as anchors for their subsequent responses, as their memory for the context may be flawed (Chen, 2010). Such an explanation related to poor learner reflections has been observed in the learning analytics literature as well (Zhou & Winne, 2012). Another explanation for inconsistency may be related to the issue of ideal self-presentation. That is, respondents may strategically alter their self-presentation during a psychological assessment in order to present themselves more favorably relative to social norms (Grieve & Elliott, 2013).

Differences between the extent and impact of response inconsistency may arise depending on the context in which the study is conducted, the characteristics of the target audience, and the sensitivity of the questions asked. For example, almost half of the participants (46%) responded inconsistently to questions about personal information such as age, gender, and educational status in an online gaming setting (Akbulut, 2015) while the degree of insufficient effort responses varied between 12% and 16% in an educational setting (Iaconelli & Wolters, 2020). In this regard, formal data collection environments may be less prone to low quality data than anonymous online environments. In terms of participants' personal characteristics, a recent empirical study suggested that respondents assigned to the careless responder class are more likely to be male, younger, unmarried, college-educated, and have higher incomes (Schneider et al., 2018). In other studies, personal interest in the research topic (Keusch, 2013) or higher academic and cognitive ability (Rosen et al., 2017) predicted better response quality. The sensitivity of the research topic has been highlighted in several papers. For example, although students gave candid responses about their course-taking patterns, their responses did not adequately reflect the truth about sensitive topics (Rosen et al., 2017). An interaction between gender and topic sensitivity was also observed in terms of the extent of inconsistent responses. Male participants, for instance, tend to underreport physical problems in order not to appear weak (Yörük Açıklık et al., 2018), whereas female participants chose to underreport their behavior when the topic is socially sensitive (Akbulut et al., 2017; Dönmez & Akbulut, 2016).

There are several methods to address low quality data resulting from inconsistent or careless responses (DeSimone et al., 2015; DeSimone & Harms, 2018). For example, direct assessment of response quality can be achieved by including validation items in a survey. Self-reported effort questions (e.g., I read all items carefully), sham items (e.g., I was born in 1979), or instructed items (e.g., Please mark *strongly disagree* for this item) can be used to weed out inconsistent responders; however, these are easily detected by participants who read all items and intentionally provide false responses. On the other hand, unobtrusive methods that are less likely to be detected by participants can be used during survey administration. That is, instead of modifying the survey with validation questions before the study, the response time or the number of consecutive and identical responses can be checked. However, determining the cutoff response time or number of consecutive identical responses to eliminate the flawed data is a tedious process (DeSimone et al., 2015). Finally, statistical methods can be implemented to deal with low quality data, such as checking for outliers or individual consistency across synonymous questions.

Discrepancies or small associations between student self-reports and objective data derived from learning management systems have received recent attention (Gasevic et al., 2017). While the use of self-reports has been the dominant approach to addressing student engagement in instructional settings (Azevedo, 2015), students may be inaccurate in calibrating their self-reported and actual behaviors in an online learning environment, and may tend to overestimate their behaviors (Winne & Jamieson-Noel, 2002). In addition to the construct of careless responding discussed above, such a discrepancy may further result from poor learner reflection or poorly reconstructed memories, such that learners' behavioral indicators in a learning system may be less biased than self-reported reflections (Zhou & Winne, 2012). Such findings have further led scholars to triangulate multiple methods to capture authentic learning processes (Azevedo, 2015; Ellis et al., 2017).

There is a tendency to benefit from learning analytic approaches in higher education in general and in open, online, and distance education in particular (Pelletier et al., 2021). In response to the widespread use of learning analytics and multiple data sources, some scholars are still cautious (Selwyn, 2020) and

have suggested asking further questions about the nature of what is really being measured, why it is really useful, and how such data relate to the learning experience (Wilson et al., 2017). Given the wide range of arguments about the reliability of self-reported data and the promise of learning analytics, we aimed to explore the alignment between self-reported and system-generated data by contextualizing the current study in an open, online, and distance education system where learning was available at scale and such data sources influence decision making in multiple dimensions.

In short, we used an unobtrusive method to identify inconsistencies between different sources of learner data in a formal open, online, and distance education system. That is, rather than adding validation items to the self-report measures, we examined response consistency by comparing different sources of self-report and learning management system (LMS) data. Based on the aforementioned literature, we hypothesized that learners' perceived intentions and actual behaviors may differ, such that their self-reported data may differ from the objective data, likely due to poor learner reflection or poorly reconstructed memories (Zhou & Winne, 2012). However, we expected that the current formal educational environment could be less prone to low-quality data than non-formal online environments such as online gaming sites (e.g., Akbulut, 2015). In line with social cognitive theory, we further hypothesized that personal and environmental factors may have played a role in the degree of response inconsistency. In this regard, we expected several variables such as participants' seniority, academic ability, gender, and satisfaction with the learning system to predict their response patterns. Finally, we hypothesized that participants' poor reflection of their actual behaviors combined with consistency-seeking needs may have led to a certain level of consistency across multiple self-report measures, in line with the concepts of anchoring and adjusting discussed above (Zhao & Linderholm, 2008). In accordance with the above literature and current hypotheses, the following research questions are investigated:

1. How similar are self-reported and LMS data?
2. What are the predictors of inconsistency between self-reported and actual use?
3. Do different sources of self-reported data (e.g., learner satisfaction, preference, and usage) support each other?

Method

Research Context

The research was conducted in an open, online, and distance education university with over two million students worldwide. The Open Education System (OES) consisted of three degree-granting colleges: The College of Open Education, The College of Economics, and The College of Business. These colleges offered a total of 60 associate or undergraduate degrees delivered entirely through open and distance learning. Students accessed courses and learning resources through an LMS. The pedagogy was primarily self-paced, while some courses include optional weekly synchronous videoconferencing sessions (i.e., live lectures). The OES allowed learners to study the learning resources online at their own time and pace, but required them to take proctored face-to-face exams to determine learner success. Applied courses within the OES also incorporated other assessment strategies such as project work. Following a multimedia approach to increase accessibility and flexibility in the learning process, a wide range of multimedia learning resources were provided online, including course books (PDF and

MP3), chapter summaries (PDF and MP3), live lectures, and practice tests. The practice tests also came in a variety of forms, including open-ended questions with extended answers, multiple-choice tests with short and extended answers, practice exams, end-of-chapter exercises, and previous semester's exam questions.

Data Collection and Cleaning

Ethics approval was granted by the institutional review board of the university. The data, then, were collected from different sources: the LMS database, satisfaction and preference questionnaires, and student information system (SIS) data for learner demographics. Learner access to resources was derived from the LMS learning analytics database. The data for each learning resource indicated whether an individual had access to the resource and the frequency of their access over the course of the semester. Self-reported data were collected for two weeks toward the end of the semester. An announcement was made on the LMS homepage, and voluntary participants who responded to the surveys were included in the current dataset.

Satisfaction and preference data came from short questionnaires. The first was a 15-item satisfaction scale developed by Open Education faculty members and used for formal and institutional research. Items were created to address student satisfaction with the open, online, and distance education system on a 5-point Likert scale ranging from 1 (*very dissatisfied*) to 5 (*very satisfied*). Exploratory factor analysis on the current dataset using maximum likelihood extraction revealed that the single-factor structure of the scale explained 77.73% of the total variance, with factor loads ranging from .84 to .92 (Cronbach's $\alpha = .98$).

In the second questionnaire, satisfaction with each of the 11 learning resources was measured with a single 5-point Likert-type question that included options such as: *This learning resource was not available in my courses* (1), *This learning resource was available but I did not use it* (2), *I used the resource but I am not satisfied* (3), *I used the resource and I am satisfied* (4), and *I used the resource and I am very satisfied* (5). This question is regularly used in institutional reports to address student usage and satisfaction.

In the third questionnaire, students were asked to select three of the 11 learning resources that they preferred the most, so that the preference score pertaining to each learning material ranged between 0 and 3. This question was deliberately used by the current research team to see the relationships between usage, satisfaction, and preference. Finally, the SIS database provided us with learner demographics such as gender, age, GPA, and current semester (i.e., 1st through 8th semesters).

Data from these resources were then combined based on unique user IDs. Duplicate responses from the same ID (if any) were removed and the most recent responses were retained. At the end of the data cleaning process, data from 20,646 students were used in the current analyses. Participants ranged in age from 17 to 75 with a mean of 32.22 ($SD: 10.6$). The average number of courses taken by participants ranged from 1 to 12, with a mean of 6.82 ($SD: 2.11$). Their semesters ranged from 1 to 8; but almost 40% of the volunteers were in their first year. The gender distribution of the participants was similar (males, 50.6%; females, 49.4%).

To identify inconsistencies between self-reported satisfaction and actual use, the following criteria were used to cross-reference the various data sources:

- IF the student response was *This learning resource was not available in my courses* BUT there was access to the learning resource, THEN an inconsistency was coded.
- IF the student response was *This learning resource was not available in my courses* AND there was no access to the learning resource, THEN consistency was coded.
- IF the student response was *This resource was available but I did not use it* BUT there was access to the resource, THEN an inconsistency was coded.
- IF the student response was *This resource was available but I did not use it* AND there was no access to the resource in question, THEN consistency was coded.
- IF the students reported usage and satisfaction/dissatisfaction (i.e., *I used* and *I am satisfied/dissatisfied*) BUT there was no access to the learning resource, THEN inconsistency was coded.
- IF the students reported usage and satisfaction/dissatisfaction AND there was access to the learning resource, THEN consistency was coded.

Accordingly, the inconsistencies between the self-reported satisfaction questionnaires and the learning analytics were determined for each of the 11 learning resources. It was also possible to calculate how many consistent (and inconsistent) answers each participant gave.

Data Analysis

Descriptive statistics were used to present self-reported and actual use, proportion of consistent responses, and preference rates. Self-reported and actual use were compared using a paired t-test. Correlations between preference rates and actual use frequencies were presented. Participants' consistency rates were presented using descriptive statistics, and predictors of consistency were examined using correlations and multiple regression. Satisfaction of actual users and non-users was compared using independent t-tests. Finally, different sources of self-reported satisfaction were investigated with further t-tests. Parametric test assumptions (e.g., normality) were checked before each analysis.

Results

Descriptive statistics of self-reported versus actual usage are summarized in Table 1. A comparison using a paired *t*-test indicated that actual usage for each of the eleven learning resources was significantly lower than self-reported usage, with a large effect size, $t(10) = 4.650$, $p < .001$, $\eta^2 = .684$. That is, students seemed to overreport their use of the learning resources. Preference was calculated by asking students to select their three favorite materials (one point each) across eleven learning resources, and the correlation between their total preference scores and their actual usage is shown in Table 1. All correlations were significant at the .001 level; however, this was likely due to the large sample size, as the correlation coefficients were quite small.

Table 1

Statistics on the Use of Learning Resources

Learning resource	Self-reported usage (%)	Actual access (%)	Access frequency	Consistent response (%)	Preference	Preference and actual use correlation (<i>r</i>)
Chapter summary (PDF)	76.0	65.8	19.66	66.3	0.56	.24*
Previous exam questions	76.0	75.3	14.04	69.5	0.39	.15*
Multiple choice questions with extended solutions	73.7	53.9	5.75	60.1	0.28	.14*
Practice exams (midterms/finals)	70.6	55.6	3.72	61.2	0.28	.2*
Open-ended questions (Q&A; PDF)	66.9	43.7	3.15	56.7	0.17	.26*
End-of-chapter exercises (multiple choice)	63.4	50.3	5.83	59.3	0.09	.26*
Coursebook (PDF)	60.3	54.0	3.33	58.1	0.47	.19*
Multiple choice questions with answer key	59.9	7.0	0.41	43.0	0.09	.07*
Audio chapter summary	43.1	11.0	0.35	58.4	0.09	.11*
Live lectures	37.0	1.7	0.05	64.1	0.15	.16*
Audio coursebook	30.4	10.5	0.24	67.9	0.05	.13*

Note. *n* = 20,646 (self-report), 18,233 (learning analytics).

* Correlations significant at the .001 level.

As shown in Table 1, the percentage of consistent responses was also calculated for each learning resource and ranged from 43% to 69.5%. If one chose to eliminate all inconsistent responses across learning resources listwise, the remaining data would look quite limited. Specifically, the number of students whose self-reported data were consistent with actual access data across all learning resources was 394 (2.2%). Table 2 shows the number of consistent responses across 11 learning resources.

Table 2

Consistent Responses Across 11 Learning Resources

Number of consistent responses	<i>f</i>	%	Cumulative %
0	126	0.7	0.7
1	397	2.2	2.9
2	647	3.5	6.4
3	936	5.1	11.6

4	1,335	7.3	18.9
5	1,793	9.8	28.7
6	2,567	14.1	42.8
7	3,018	16.6	59.3
8	3,084	16.9	76.3
9	2,531	13.9	90.1
10	1,405	7.7	97.8
11	394	2.2	100.0
Total	18,233	100.0	

The number of consistent responses per student ranged from 0 to 11 ($M = 6.65$, $SD = 2.38$) with a relatively normal distribution (skewness = -0.51; kurtosis = -0.26). A gender comparison revealed that females ($M = 6.81$; $SD = 2.27$) were more consistent than males ($M = 6.47$; $SD = 2.48$) with a statistically significant difference but with a small effect size, $t(18,221) = 9.73$, $p < .001$, $\eta^2 = .005$. Student consistency was positively correlated with the number of visits to the LMS ($r = .146$; $p < .001$), GPA ($r = .097$; $p < .001$), student satisfaction ($r = .02$; $p < .006$), and age ($r = .023$; $p < .002$), while it was negatively correlated with students' current semester ($r = -.072$; $p < .001$). However, these variables explained only 4% of the total variance in response consistency, $R = .201$; $R^2 = .04$, $F(6; 8,475) = 59.18$, $p < .001$. Model coefficients and t -values are shown in Table 3. It should be noted that age was not a statistically significant predictor when entered into the model with the other predictors in the current study.

Table 3

Predictors of Inconsistency

Predictors in the model	Unstandardized coefficient		Standardized coefficient	t	p
	B	Std. error	Beta		
(Constant)	6.371	0.158		40.278	<.001
GPA	0.296	0.036	0.092	8.324	<.001
Age	0.001	0.003	0.003	0.244	.807
Semester	-0.073	0.012	-0.064	-5.888	<.001
Number of visits to the system	0.012	0.001	0.120	10.855	<.001
Gender	-0.425	0.053	-0.088	-7.982	<.001
Self-reported satisfaction	0.082	0.025	0.035	3.260	.001

Note. Dependent variable: consistency.

Learning analytics data and self-reported satisfaction scores were also used to compare the average satisfaction scores of users who actually visited a particular learning resource with the average satisfaction scores of non-users (who never visited a particular learning resource). With the exception of PDF and audio coursebooks, the satisfaction scores of users were slightly higher than those of non-

users, as summarized in Table 4. However, the means of both groups were already high, as indicated by a negatively skewed and leptokurtic distribution (skewness = -1.21; kurtosis = 1.05). In addition, the effect sizes associated with these comparisons were very small. Accordingly, the number of visits to each learning resource did not show substantial correlations with the average satisfaction scores. More specifically, the actual use of each learning resource could explain a trivial amount of the variance in satisfaction scores, — $R = .08$; $R^2 = .007$; $F(11; 18,221) = 11.47$; $p < .001$. Preference rates pertaining to each learning resource and satisfaction scores were not substantially related either, — $R = .14$; $R^2 = .02$; $F(11; 20,634) = 38.67$; $p < .001$.

Table 4

Satisfaction of Users and Non-Users

Learning resource	Usage	<i>n</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i>	η^2
Chapter summary (PDF)	No	6,235	3.98	1.02	-4.88	18,231	<.001	.001
	Yes	11,998	4.05	0.97				
Previous exam questions	No	4,510	3.97	1.02	-4.53	18,231	<.001	.001
	Yes	13,723	4.04	0.97				
Multiple choice questions with extended solutions	No	8,411	3.96	1.02	-7.79	18,231	<.001	.003
	Yes	9,822	4.08	0.95				
Practice exams (midterms/finals)	No	8,087	3.98	1.01	-6.01	18,231	<.001	.002
	Yes	10,146	4.06	0.96				
Open-ended questions (Q&A; PDF)	No	10,263	3.97	1.02	-9.29	18,231	<.001	.005
	Yes	7,970	4.10	0.93				
End-of-chapter exercises (multiple choice)	No	9,054	3.97	1.01	-7.02	18,231	<.001	.003
	Yes	9,179	4.08	0.96				
Coursebook (PDF)	No	8,388	4.03	0.99	0.99	18,231	.323	<.001
	Yes	9,845	4.02	0.97				
Multiple choice questions with answer key	No	16,952	4.02	0.99	-3.09	18,231	.002	.001
	Yes	1,281	4.11	0.90				
Audio chapter summary	No	16,226	4.02	0.99	-3.72	18,231	<.001	.001
	Yes	2,007	4.10	0.91				
Live lectures	No	17,917	4.02	0.99	-3.67	18,231	<.001	.001
	Yes	316	4.23	0.82				
Audio coursebook	No	16,325	4.03	0.99	0.45	18,231	.655	<.001
	Yes	1,908	4.02	0.96				

Through the aforementioned analyses, we suggested an inconsistency between the objective data derived from the open, online, and distance education system and the subjective data (i.e., self-reports).

In addition, it was not possible to maintain a substantial relationship between the satisfaction, preference and the actual use. However, the validation of the 15-item satisfaction scale with self-reported usage was somewhat successful. Specifically, students who reported use and satisfaction (i.e., *I used the resource and I am satisfied/very satisfied*) were compared with those who reported use but dissatisfaction (i.e., *I used the resource but I am not satisfied*). Almost all comparisons resulted in large effect sizes, as summarized in Table 5. That is, two separate self-report measures of satisfaction were somewhat consistent.

Table 5

Consistency Between the Two Separate Measures of Satisfaction

Learning resource	I am	<i>n</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>P</i>	η^2
Chapter summary (PDF)	Dissatisfied	1,788	2.93	1.19	-56.82	15,686	<.001	.171
	Satisfied	13,900	4.20	0.84				
Previous exam questions	Dissatisfied	1,142	2.84	1.24	-46.58	15,696	<.001	.121
	Satisfied	14,556	4.15	0.88				
Multiple choice questions with extended solutions	Dissatisfied	1,266	2.80	1.21	-52.27	15,211	<.001	.152
	Satisfied	13,947	4.18	0.86				
Practice exams (midterms/finals)	Dissatisfied	1,328	2.89	1.21	-49.38	14,566	<.001	.143
	Satisfied	13,240	4.18	0.87				
Open-ended questions (Q&A; PDF)	Dissatisfied	1,494	2.90	1.22	-52.67	13,803	<.001	.167
	Satisfied	12,311	4.20	0.85				
End-of-chapter exercises (multiple choice)	Dissatisfied	1,487	2.94	1.22	-50.66	13,090	<.001	.164
	Satisfied	11,605	4.21	0.86				
Coursebook (PDF)	Dissatisfied	2,180	3.04	1.17	-53.24	12,451	<.001	.185
	Satisfied	10,273	4.20	0.87				
Multiple choice questions with answer key	Dissatisfied	1,310	2.93	1.24	-47.62	12,363	<.001	.155
	Satisfied	11,055	4.20	0.87				
Audio chapter summary	Dissatisfied	1,676	3.00	1.23	-51.06	8,896	<.001	.227
	Satisfied	7,222	4.29	0.84				
Live lectures	Dissatisfied	1,416	2.97	1.24	-47.44	7,627	<.001	.228
	Satisfied	6,213	4.29	0.87				
Audio coursebook	Dissatisfied	1,531	3.02	1.25	-44.3	6,266	<.001	.238
	Satisfied	4,737	4.31	0.89				

Discussion

The current research signaled a discrepancy between objective student behavior (i.e., tracking data through digital footprints) derived from the learning management system and subjective data (i.e., self-reports), which supports the findings of empirical studies in the literature (Gasevic et al., 2017; Zhou & Winne, 2012). More specifically, students overreported their use. This could be due to either insufficient motivation to respond, intentional falsification (i.e., faking), or poor recall of learning experiences by students. While the source of such discrepancies should be explored through further research, scholars may choose to use a combination of multiple methods to better reflect the processes used during learning (Azevedo, 2015; Ellis et al., 2017). Learner metacognition may be specifically considered as a covariate when making decisions about inconsistency, as either poor learner reflection or poorly reconstructed memories may have resulted in low-quality data (Zhou & Winne, 2012).

Inconsistency was observed even though the content was not culturally sensitive and even though the setting was a formal learning environment. Furthermore, learners' gender, age (Schneider et al., 2018), and their academic ability (Rosen et al., 2017) predicted consistency, as expected. While the degree of consistency varied across learning materials, both actual use and learner satisfaction were associated with the degree of consistency. In this regard, when learning materials are more satisfying and useful, there seems to be a greater match between what learners say and what the system data provides. However, we do not know about the perceived quality and usefulness of the learning resources as rated by the learners. In this regard, further research could include the perceived usefulness and quality of learning materials as variables of interest.

Students' current semester was negatively correlated with consistency. We speculated that because students were asked to respond to multiple online surveys over the course of their undergraduate studies, survey fatigue may have led to an overdose of research participation and thus higher levels of careless responding. While there were slight differences between actual users and non-users in terms of satisfaction, the overall satisfaction average was very high. In addition, the number of visits to each learning resource was not strongly correlated with satisfaction scores. That is, even learners who did not use the system were satisfied with it. This was considered quite problematic, since it may not be right to make policy decisions based on students' judgments about a system they do not actually use. Similarly, students' preferences and actual use were correlated due to the large sample size, but the coefficients were quite small. Thus, their self-reported preferences did not show a substantial relationship with their actual usage patterns. Several empirical studies have often used student satisfaction (e.g., Alqurashi, 2019; So & Brush, 2008; Wu et al., 2010), intention to use the online learning systems (e.g., Chao, 2019), or learner preferences (e.g., Rhode, 2009; Watson et al., 2017) to evaluate online learning environments. However, the current findings suggested that objective system or performance data should be considered in addition to self-reports in order to draw more robust implications regarding the accountability of online learning systems. In addition, current LMS data is primarily limited to the presence and frequency of access to specific learning resources. Additional objective data sources and variables related to online learning experiences need to be integrated to support or refute current hypotheses.

While we were able to identify some of the predictors of inconsistencies between self-report and LMS data, we were only able to explain a very small percentage of the variability. In this regard, alternative variables from the field of learning analytics can be integrated. On the other hand, the consistency between two sources of subjective data addressing the same construct (i.e., learner satisfaction) was

strong. While the inclusion of such validation items and scales in the research design has been considered as a method to directly assess response quality (DeSimone & Harms, 2018), this was not the case between self-reported and LMS data. That is, our findings suggested that two self-reported data sources may sometimes be compatible with each other, but both may be at odds with the actual usage data. In this regard, unobtrusive methods may be more effective at eliminating low-quality data than integrating validation items. To test this speculation, future researchers could compare the effects of obtrusive and unobtrusive validation methods on multiple groups. In addition, we did not record participants' survey response times, which may be considered as a covariate in further studies.

A critical implication of the current study is to consider the unreliability of self-report data, which is commonly used in educational research to inform policy decisions. In addition to using alternative data collection tools, we need to look for more objective and direct measures. We have tended to focus a great deal on the reliability of measures in general, and the internal consistency of items in particular, to the detriment of validity (Steger et al., 2022). The survey itself was not the only source of measurement error observed in the current study. Participants can also be a critical source of erroneous data. In addition to attitudes and reflections, which may be over- or underreported depending on the sensitivity of the issue, we need to use actual performance data as well. For example, while years of self-report research have emphasized that men have an advantage in technical competence, systematic analyses using performance-based measures have found that the opposite may be true (Borgonovi et al., 2023; Siddiq & Scherer, 2019). These limitations, combined with the implications of the current study, support calls from eminent scholars for robust intervention research that should include sound measures and variables to address relevant instructional technology problems (Hsieh et al., 2005; Reeves & Lin, 2020; Ross & Morrison, 2008). These findings also suggested that strategic planning decisions that guide short-, medium-, and long-term goals can be based not only on self-reported data, but also on learning analytics data available in most LMSs. We recognize the potential of the current findings to unsettle the social science community at large, where thousands of self-report studies are conducted each year. On the other hand, if we do not integrate alternative and more objective data sources into more robust designs, it is likely that the replication crisis will continue.

Concluding Details

The following are details about specific aspects of how this research was conducted. First, this research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The authors declare that there are no conflicts of interest related to this article. Data will be made available upon reasonable request. Finally, our research proposal was approved by the Institutional Review Board of Anadolu University (March 28, 2023, No: 33/63).

References

- Akbulut, Y. (2015). Predictors of inconsistent responding in web surveys. *Internet Research*, 25(1), 131–147. <https://doi.org/10.1108/IntR-01-2014-0017>
- Akbulut, Y., Dönmez, O., & Dursun, Ö. Ö. (2017). Cyberloafing and social desirability bias among students and employees. *Computers in Human Behavior*, 72, 87–95. <https://doi.org/10.1016/j.chb.2017.02.043>
- Alqurashi, E. (2019). Predicting student satisfaction and perceived learning within online learning environments. *Distance Education*, 40(1), 133–148. <https://doi.org/10.1080/01587919.2018.1553562>
- Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist*, 50(1), 84–94. <https://doi.org/10.1080/00461520.2015.1004069>
- Baltar, F., & Brunet, I. (2012). Social research 2.0: Virtual snowball sampling method using Facebook. *Internet Research*, 22(1), 57–74. <https://doi.org/10.1108/10662241211199960>
- Bandura, A. (1977). *Social learning theory*. Prentice Hall.
- Borgonovi, F., Ferrara, A., & Piacentini, M. (2023). From asking to observing. Behavioural measures of socio-emotional and motivational skills in large-scale assessments. *Social Science Research*, 112, 102874. <https://doi.org/10.1016/j.ssresearch.2023.102874>
- Bozkurt, A., Akgun-Özbek, E., Onrat-Yilmazer, S., Erdoğan, E., Uçar, H., Güler, E., Sezgin, S., Karadeniz, A., Sen, N., Göksel-Canbek, N., Dinçer, G. D., Arı, S., & Aydın, C. H. (2015). Trends in distance education research: A content analysis of journals 2009–2013. *International Review of Research in Open and Distributed Learning*, 16(1), 330–363. <http://dx.doi.org/10.19173/irrodl.v16i1.1953>
- Castro, R. (2013). Inconsistent respondents and sensitive questions. *Field Methods*, 25(3), 283–298. <https://doi.org/10.1177/1525822x12466988>
- Chao, C. M. (2019). Factors determining the behavioral intention to use mobile learning: An application and extension of the UTAUT model. *Frontiers in Psychology*, 10, 1652. <https://doi.org/10.3389/fpsyg.2019.01652>
- Chen, P. H. (2010). *Item order effects on attitude measures* (Publication No. 778) [Doctoral dissertation, University of Denver]. Electronic Theses and Dissertations. <https://digitalcommons.du.edu/etd/778>
- Chesney, T., & Penny, K. (2013). The impact of repeated lying on survey results. *SAGE Open*, 3(1), 1–9. <https://doi.org/10.1177/2158244012472345>
- DeSimone, J. A., Harms, P. D., & DeSimone, A. J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior*, 36, 171–181. <https://doi.org/10.1002/job.1962>

- DeSimone, J. A., & Harms, P. D. (2018). Dirty data: The effects of screening respondents who provide low-quality data in survey research. *Journal of Business and Psychology*, 33, 559–577. <https://doi.org/10.1007/s10869-017-9514-9>
- Dönmez, O., & Akbulut, Y. (2016). Siber zorbalık çalışmalarında sosyal beğenirlik etmeni [Social desirability bias in cyberbullying research]. *Eğitim Teknolojisi Kuram ve Uygulama*, 6(2), 1–18. <https://dergipark.org.tr/tr/pub/etku/issue/24420/258838>
- Ellis, R. A., Han, F., & Pardo, A. (2017). Improving learning analytics—Combining observational and self-report data on student learning. *Journal of Educational Technology & Society*, 20(3), 158–169. <https://www.jstor.org/stable/26196127>
- Evans, J. R., & Mathur, A. (2005). The value of online surveys. *Internet Research*, 15(2), 195–219. <https://doi.org/10.1108/10662240510590360>
- Gasevic, D., Jovanovic, J., Pardo, A., & Dawson, S. (2017). Detecting learning strategies with analytics: Links with self-reported measures and academic performance. *Journal of Learning Analytics*, 4(2), 113–128. <https://doi.org/10.18608/jla.2017.42.10>
- Gregori, A., & Baltar, F. (2013). ‘Ready to complete the survey on Facebook’: Web 2.0 as a research tool in business studies. *International Journal of Market Research*, 55(1), 131–148. <https://doi.org/10.2501/ijmr-2013-010>
- Grieve, R., & Elliott, J. (2013). Cyberfaking: I can, so I will? Intentions to fake in online psychological testing. *Cyberpsychology, Behavior, and Social Networking*, 16(5), 364–369. <https://doi.org/10.1089/cyber.2012.0271>
- Hsieh, P., Acee, T., Chung, W., Hsieh, Y., Kim, H., Thomas, G., Levin, J. R., & Robinson, D. H. (2005). Is educational intervention research on the decline? *Journal of Educational Psychology*, 97(4), 523–529. <https://doi.org/10.1037/0022-0663.97.4.523>
- Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, 100(3), 828–845. <https://doi.org/10.1037/a0038510>
- Iaconelli, R., & Wolters, C. A. (2020). Insufficient effort responding in surveys assessing self-regulated learning: Nuisance or fatal flaw? *Frontline Learning Research*, 8(3), 104–125. <https://doi.org/10.14786/flr.v8i3.521>
- Kara Aydemir, A. G., & Can, G. (2019). Educational technology research trends in Turkey from a critical perspective: An analysis of postgraduate theses. *British Journal of Educational Technology*, 50(3), 1087–1103. <https://doi.org/10.1111/bjet.12780>
- Keusch, F. (2013). The role of topic interest and topic salience in online panel web surveys. *International Journal of Market Research*, 55(1), 59–80. <https://doi.org/10.2501/ijmr-2013-007>

- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3), 213–236.
<https://doi.org/10.1002/acp.2350050305>
- Küçük, S., Aydemir, M., Yildirim, G., Arpacik, O., & Goktas, Y. (2013). Educational technology research trends in Turkey from 1990 to 2011. *Computers & Education*, 68, 42–50.
<https://doi.org/10.1016/j.compedu.2013.04.016>
- Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, 48, 61–83.
<https://doi.org/10.1016/j.jrp.2013.09.008>
- Mishra, P., Koehler, M. J., & Kereluik, K. (2009). Looking back to the future of educational technology. *TechTrends*, 53(5), 48–53. <https://doi.org/10.1007/s11528-009-0325-3>
- Pekrun, R. (2020). Commentary: Self-report is indispensable to assess students' learning. *Frontline Learning Research*, 8(3), 185–193. <https://doi.org/10.14786/flr.v8i3.637>
- Pelletier, K., Brown, M., Brooks, D. C., McCormack, M., Reeves, J., Arbino, N., Bozkurt, A., Crawford, S., Czerniewicz, L., Gibson, R., Linder, K., Mason, J., & Mondelli, V. (2021). 2021 EDUCAUSE Horizon report teaching and learning edition. EDUCAUSE.
<https://www.learntechlib.org/p/219489/>
- Reeves, T. C., & Lin, L. (2020). The research we have is not the research we need. *Educational Technology Research and Development*, 68, 1991–2001. <https://doi.org/10.1007/s11423-020-09811-3>
- Rhode, J. (2009). Interaction equivalency in self-paced online learning environments: An exploration of learner preferences. *The International Review of Research in Open and Distributed Learning*, 10(1). <https://doi.org/10.19173/irrodl.v10i1.603>
- Rosen, J. A., Porter, S. R., & Rogers, J. (2017). Understanding student self-reports of academic performance and course-taking behavior. *AERA Open*, 3(2).
<https://doi.org/10.1177/2332858417711427>
- Ross, S. M., & Morrison, G. R. (2008). Research on instructional strategies. In M. Spector, M. D. Merrill, J. V. Merrienboer, & M. Driscoll (Eds.). *Handbook of research on educational communications and technology* (3rd ed., pp. 719–730). Routledge.
https://doi.org/10.1007/978-1-4614-3185-5_3
- Schneider, S., May, M., & Stone, A. A. (2018). Careless responding in Internet-based quality of life assessments. *Quality of Life Research*, 27, 1077–1088. <https://doi.org/10.1007/s11136-017-1767-2>
- Selwyn, N. (2020). Re-imagining 'learning analytics'... a case for starting again? *The Internet and Higher Education*, 46, 100745. <https://doi.org/10.1016/j.iheduc.2020.100745>

- Siddiq, F., & Scherer, R. (2019). Is there a gender gap? A meta-analysis of the gender differences in students' ICT literacy. *Educational Research Review*, 27, 205–217. <https://doi.org/10.1016/j.edurev.2019.03.007>
- So, H. J., & Brush, T. A. (2008). Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: Relationships and critical factors. *Computers & Education*, 51(1), 318–336. <https://doi.org/10.1016/j.compedu.2007.05.009>
- Steger, D., Jankowsky, K., Schroeders, U., & Wilhelm, O. (2022). The road to hell is paved with good intentions: How common practices in scale construction hurt validity. *Assessment*. <https://doi.org/10.1177/10731911221124846>
- Watson, S. L., Watson, W. R., Yu, J. H., Alamri, H., & Mueller, C. (2017). Learner profiles of attitudinal learning in a MOOC: An explanatory sequential mixed methods study. *Computers & Education*, 114, 274–285. <https://doi.org/10.1016/j.compedu.2017.07.005>
- Wilson, A., Watson, C., Thompson, T. L., Drew, V., & Doyle, S. (2017). Learning analytics: Challenges and limitations. *Teaching in Higher Education*, 22(8), 991–1007. <https://doi.org/10.1080/13562517.2017.1332026>
- Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology*, 27(4), 551–572. [https://doi.org/10.1016/S0361-476X\(02\)00006-1](https://doi.org/10.1016/S0361-476X(02)00006-1)
- Wu, J. H., Tennyson, R. D., & Hsia, T. L. (2010). A study of student satisfaction in a blended e-learning system environment. *Computers & Education*, 55(1), 155–164. <https://doi.org/10.1016/j.compedu.2009.12.012>
- Yörük Açıklı, B., Turhan, U., & Akbulut, Y. (2018). Effect of multitasking on simulator sickness and performance in 3D aerodrome control training. *Simulation & Gaming*, 49(1), 27–49. <https://doi.org/10.1177/1046878117750417>
- Zhao, Q., & Linderholm, T. (2008). Adult metacomprehension: Judgment processes and accuracy constraints. *Educational Psychology Review*, 20, 191–206. <https://doi.org/10.1007/s10648-008-9073-8>
- Zhou, M., & Winne, P. H. (2012). Modeling academic achievement by self-reported versus traced goal orientation. *Learning and Instruction*, 22(6), 413–419. <https://doi.org/10.1016/j.learninstruc.2012.03.004>
- Zhu, M., Sari, A. R., & Lee, M. M. (2020). A comprehensive systematic review of MOOC research: Research techniques, topics, and trends from 2009 to 2019. *Educational Technology Research and Development*, 68, 1685–1710. <https://doi.org/10.1007/s11423-020-09798-x>

