Iranian Journal of Applied Linguistics (IJAL)

Vol. 24, No. 2, September 2021, 57-81

A Corpus-driven Study on Identity Option Depicted in the Academic Essays Written by University Students: A Transformative Mixed-Method Design

Fatemeh Esmaeeli

PhD candidate in TEFL, Shiraz University, Shiraz, Iran

Naser Rashidi*

Professor, Shiraz University, Shiraz, Iran

Abstract

Authorial identity is a notion refers to the way an author expresses their sense of self, employs their agency, and develops their academic discourse. Following a transformative mixed-method design, and by adopting Fairclough's (1992) *Discourse as Text*, *Interaction*, and *Context* framework, authorial identity option was investigated among university students. To this end, 540 essays written by university students were examined in terms of identity-related factors, i.e. social factors", "agency", "knowledge & discourses", "gender", "group", "education", as well as "disciplinary group" and "presentation or non-presentation of authorial identity". The results of the study showed that authorial identity is a complex process restructured through the process of negotiation with various individual, discoursal, sociocultural, and demographic characteristics. It was concluded that those identity-related factors may develop individuals to consider self-legitimacy in employing authorial identity, though it is also rooted in epistemological preferences of their disciplines.

Keywords: Identity option, Authorial identity, identity indicators, Identity variables, Self-legitimacy, Epistemological preferences

Corresponding author: Department of Linguistics & Foreign Languages Shiraz University, Shiraz, Iran *Email:* naser.rashidi@shirazu.ac.ir

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

1. Introduction

The past three decades show a new paradigm in the field of education in general and language teaching/learning in particular, in which identity has been underscored as the main factor for achievement or underachievement among the learners (e.g., Gee, 2004; Norton, 2013; Norton Peirce, 1995; Toohey, Manyak, & Day, 2007). Furthermore, achievement and underachievement patterns are highly associated with identity negotiation, investment, and affirmation (Cummins et al., 2015). The importance of identity for academic outcomes is in such a way that it is called "identities of competence" by Manyak (2004). Identity of competency, according to Stone and Harrt (2019, p. 70), is "an analytical frame that conceptualizes competency as an emergent, situated, and implicitly inhabited social identity". An identity of competence is, in fact, "an example of an implicit form of social identity that is linked to a local community's socially shared but tacit category for competency (which) represents the values and expectations for how to be, act, or feel like a competent member of a community" (Stone & Hart, 2019, p. 70). This notion is "implicitly valued and highly desired because it brings recognition and status (rights/power) for regulating the behaviors of self and others as well as indicates a willingness to persist in complex learning" (Stone & Hart, 2019, p. 71).

Moreover, the related literature illustrates that societal power and its influence on identity negotiation patterns in educational settings are the main causes of students' underachievement, especially the ones from lower social communities with discrimination and limited educational and professional opportunities (Cummins et al., 2015). In addressing underachievement among students from marginalized communities, it is necessary to implement pedagogies to enhance learners' identities of competence. Recently, the notion of "identity text" introduced by some researchers denotes the connection among identity affirmation, power relations, and literacy engagement (Cummins, 2004; Cummins & Early, 2011). Identity text, in effect, emphasizes the background knowledge brought by the students into the class as a cognitive and identity affirmation tool to enhance students' literacy engagement (Cummins, 2015; Cummins & Early, 2011; Walker, 2014).

In this regard, Cummins (2001) and Cummins and Early (2011) argue that "creative writing and other forms of cultural production or performance (e.g., art, drama, video creation, etc.) represent expressions of identity, projection of identity into new social spheres, and re-

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

creation of identity as a result of feedback from and dialogue with multiple audiences" (cited in Cummins et al., 2015, p. 557). Accordingly, identity text to Cummins et al., (2015) refers to the students' identities investment which can be oral, written, visual, or multimodal. Identity investment is a notion in sociocultural approaches which deals with the multiple, changing, and contradictory identities depicted especially by learners. Therefore, the students' identities investment can be approached as the identities (authorial or non-authorial) represented by students in their writing tasks.

An important point to be considered is that students' identities sharing with others including teachers, peers, family members may result in identities' affirmations, which may positively influence their literacy achievement. Identities' affirmation is especially significant for students from the minority communities, whose language, culture, religions, and gender are different from the majority group. The foundation of language and identity was developed mainly by Norton (1997), who in defining the relationship between language and identity refers to numerous theories such as the ones by Bourdieu (1977), West (1992), and Weedon (1987). Identity, in effect, is a dialogical relationship between being and becoming in which individuals' past, present and future are negotiated (Giampapa, 2004).

Identity, in effect, is at the heart of anything individuals do, from learning to working. For example, identity seems to play a mediating role in educational settings, which affects learners' performances. The importance of identity is even more critical in higher education, and especially in academic essays written by learners since it may affect their achievement or underachievement. In this regard, Ivanic (1998) argues that learners by writing do not only convey content but also represent their identity. She adds that one of the reasons that makes writing so challenging for the students is rooted in the fact that they have problems in portraying self in their writing. Ivanič (1998), in effect, views writing as "an act of identity in which people align themselves with socioculturally-shaped possibilities of selfhood, playing their part in reproducing or challenging dominant practices and discourses, and the values, beliefs and interests which they embody" (p. 31).

The above discussion clarifies that identity is at the center of any research, ranging from research on the micro factors including individual features (e.g. motivation) to the macro ones including sociocultural or political contexts. In fact, identity is the cornerstone of learning or teaching. It is highly related to motivation and learners' subjectivity. The main cause of learners' investment to learn is the notion of "identity". Meanwhile, identity is mediating between an

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

individual's past, present and future. It may enable or disable learners to see themselves as legitimate speakers, which directly or indirectly influences their capabilities.

Review of literature also shows that academic identity is a controversial concept with different definitions and perspectives. Whereas, it means professional development in academic settings for Winters (2009), Welch and Hodges (1997, p.37) view it as "the personal commitment to a standard of excellence, the willingness to persist in the challenge, struggle, excitement and disappointment intrinsic in the learning process". Meanwhile, there are various terminologies in referring to academic identity including "educational identity" (Hejazi et al., 2010), "student identity", and "school-affiliated identity" (Hawkins, 2005) which project various dimensions of academic identity.

Writing is apparently a locus of authorial identity portrayal by university students. This issue seems to be challenging, as various differences were detected in the students' essays by the researchers. So, it can be hypothesized that the problem may be due to various sociodemographic backgrounds among the students. Therefore, an obvious, yet perplexing, problem may be linked to students' unfamiliarity with how to portray their authorial identity.

The aforementioned issues motivated the researchers to explore identity manifestation in the academic essays written by the students as their assignments by adopting Fairclough's (1992) *Discourse as Text, Interaction,* and *Context* framework. In effect, the main thrust of the study was to see whether there is any relationship between academic identity of university students and their writing performances. To this end, we utilized two groups of data, i.e. The British Academic Written English (BAWE) corpus (216 samples) and the essays written by Iranian and Iraqi students written as part of the assessment for their courses (324 samples). In line with the purposes of the study, the following questions were addressed:

1. Is there any relationship between students' identity-related variables and their authorial identity option depicted in their academic essays?

2. Is there any relationship between students' writing performance and their identity option?

3. Can disciplinary group be considered as a factor affecting identity option taken by the students?

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

2. Method

2.1. Design of the Study

Following a transformative mixed-method design study, two groups of data were utilized, namely, university students' essays studying in a foreign context and British Academic Written English (BAWE) corpus.

2.2. Materials

In conducting the present study, we used two groups of essays. The first group was the essays written by 324 non-native university students (116 Iranian students and 208 Iraqi students) as part of the assessment of their courses who were studying in an EFL context wherein English was considered as a foreign language. Table 1 shows the sociodemographic information of the essays written by EFL students.

Table 1. *Sociodemographic Information of the Included Essays* (EFL Context)

		Frequency	Percent
Age	=18-25	284	87
	=26-30	20	6
	=31-34	5	2
	=35-40	15	5
Gender	=female	130	40.1
	=male	194	59.9
Native Language	=English	0	0
	=non-English	324	100
Disciplinary groups	=Arts & Humanities	259	79.9
	=Social Sciences	23	7.1
	=physical Sciences	42	13
Education	=UKA	0	0
	=partly in UKA & partly Overseas	0	0
	=solely Overseas	324	100
Level	=first year	243	75
	=second year	71	21.9
	=third year	10	3.1
	=fourth year (master)	0	12.9
Grade	=M (60%-69%)	82	25.3
	=D (70%-100%)	242	74.7

As Table 1 shows 324 essays written by the EFL university students were selected from the two universities of Shiraz (36.5%) and Samara (63.5%). The essays had been written as the requirements of their courses by the students. Those essays were gathered from their

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

instructors. However, we had access to the British Academic Written English (BAWE) corpus, we also included the essays from the two universities of Shiraz and Samara in order to present more dependable results by adding an EFL sociocultural context (Iran & Iraq) which is different from the UK. The included essays had been written by Iraqi and Iranian EFL students from 2019 to 2022.

British Academic Written English (BAWE) corpus, on the other hand, comprises the second source of data. The rationale for including this corpus was that it included a great number of learners' academic essays, which were authentically written in order that their performance on the courses are measured. Moreover, including the corpus enabled the researchers to have access to authentic contexts and situations. In fact, the corpus enabled us to conduct the research in relation to a number of ecologically identity-related features such as age, gender, level of education, socioeconomic background, place and time of the communicative event (writing essays).

The BAWE corpus was developed at the Universities of Warwick, Reading and Oxford Brookes, under the directorship of Hilary Nesi and Sheena Gardner (formerly of the Centre for Applied Linguistics [previously called CELTE], Warwick), Paul Thompson (Department of Applied Linguistics, Reading) and Paul Wickens (Westminster Institute of Education, Oxford Brookes), with funding from the ESRC.

However, the corpus consists of writings with various genres (e.g. case study, critique, exercise, essay, proposal, etc.), we included only the writings with the genre of "essay" from the three disciplinary groups of "Arts & Humanities", "social sciences", and "life sciences". It should be noted that essays were written by students at different levels of university (1st year, 2nd year, or 3rd year of undergraduate or graduate programs). Having included the essays from students with different levels enabled us to see if there are any differences in identity option due to the levels at which the essays were written. Likewise, 216 essays were selected from the corpus. Table 2 sheds light on the sociodemographic information of the included corpus-based essays.

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

Table 2. *Sociodemographic Information of the Included Essays from the BAWE*

		Frequency	Percent
Age	=16-20	58	26.9
	=21-25	92	42.5
	=26-30	34	15.7
	=31-35	32	14.9
Gender	=female	117	54.1
	=male	99	43.9
Native Language	=English	148	68.5
	=non-English	59	27.3
Disciplinary groups	=Arts & Humanities	78	36.1
	=Social Sciences	89	41.2
	=Life Sciences	45	20.8
Education	=UKA	125	57.8
	=partly in UKA & partly Overseas	16	7.4
	=solely Overseas	66	30.5
Level	=first year	116	53.7
	=second year	56	25.9
	=third year	16	7.4
	=fourth year (master)	28	13
Grade	=M (60%-69%)	113	52.3
	=D (70%-100%)	94	43.5

N.B. The observed differences between the total number of the essays and the provided sociodemographic information in the table are that some sociodemographic information was missing in a few essays(9 essays lacked the information regarding native language, 4 essays lacked the data of the disciplinary group, 9 essays lacked the data of education, and 9 essays lacked the grade information)

As Table 2 shows, the essays written by the students with the ages between 16 and 35, include both genders from various university levels (both undergraduate and master students) with school education either in UK, overseas, or partly in UK and out of UK. It is worth

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

mentioning that the BAWE corpus was funded by the ESRC from 2004 to 2007 at the universities of Warwick, Reading and Oxford Brookes.

2.3. Underpinnings of the Study

Having followed the tenets of sociocultural theory, we adopted Fairclough's (1989) *Discourse as Text, Interaction, and Context* framework, in which sociocultural context is fundamental. This framework enabled us to consider three layers in dealing with university students' authorial identity (Figure 1). The first layer shows us the essays we gathered (the data). The second layer, in effect, demonstrates the audience which in this case are the examiners who read and scored the essays. Layer 3 is the sociocultural context of the five universities, wherein the essays are written (the three universities of Warwick, Reading and Oxford Brookes between 2004 and 2007, and the two universities of Shiraz and Samara between 2019 and 2022). The sociocultural context (layer 3) also involves the disciplinary groups of various essays. The sociocultural context leads the students to write based on the conventions of institutional policies. All the three layers are interacting constantly and lead the students to write the required essays.

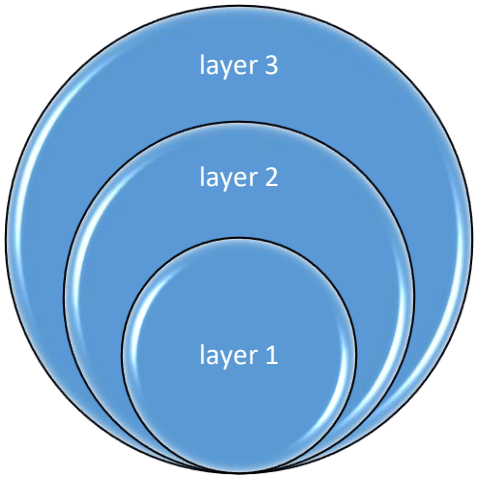

Figure 1. Conceptual Model Adapted from Fairclough's (1989) Discourse as Text, Interaction, and context framework

2.4. Data Collection and Analysis Procedures

The sources of our data were the BAWE corpus and the essays written by EFL university students. It should be noted that the essays were written by the students to be assessed by their instructors and was related to the courses they had taken. Using the Sketch Engine of the corpus, we downloaded the essay genres in three fields of "Arts & Humanities", "social

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

sciences", and "life sciences", out of which 216 essays were selected randomly. Moreover, the 324 essays written by Iranian and Iraqi students related to their course (General English & Reading 1) were gathered from the lecturers of those courses. It should be noted that the two groups of data were tagged with the same information as the corpus. In fact, each essay included the sociodemographic information related to the author (the students who had written those essays), including their age, native language, level (undergraduate 1st year, undergraduate 2nd year, undergraduate 3rd year, graduate), disciplinary group ("Arts & Humanities", "Social Sciences", "Life Sciences"), education (solely in UK, Partly in UK and partly out of the UK, completely overseas), grades (the quality of essays assessed by the instructors of those courses, i.e. Merit (M):50%-69%, Distinction (D): 70%-100%) and group (native speakers of English, non-native speakers of English). The EFL students' essays, which had been previously graded out of 20 by their instructors, were transformed to percentages and classified into the two groups of D and M.

Beside the above-mentioned data which were extracted from the essays, coded and put into SPSS, we went through a meticulous text analysis carried out in MAXQDA. In effect, we developed a sketch based on which we analyzed the texts (writings) and extract any representation of first singular or plural pronouns (I, We, Me, Us, My, Our).

In addition to the above variables tagged to various essays, and various authorial identity options located in the documents, we tried to classify the documents in terms of social actors (audiences), agency, knowledge and discourses, and social and contextual factors.

Particularly, the above added variables helped us to clarify various dimensions of the sociocultural settings which were fundamental for our developed framework (layer 3). In clarifying social actors in the essays, we classified the documents in terms of the class, instructor, and presumably classmates, which were recognized through the class number and the name of the subject (course) determined and coded for each essay. For social and contextual factors, we were dealing with educational settings including university, school, college, etc. which explicated sociocultural conditions and academic-non-academic encounters of the authors. Agency was recognized in terms of students' interpretation and understanding of academic subjects represented in their essays as their particular ideas regarding the subject they were talking about. Knowledge and discourses were recognized through the disciplinary groups, native language, additional languages of the authors, and the discussion presented by students in their essays.

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

In answering the three raised questions, we ran several binary logistic regressions in order to examine if authorial identity option represented by the students in their essays significantly predicted their writing performance, and also to see if sociodemographic information of the essays influence identity option projection in the essays. In addition, another binary logistic regression test was utilized to explore if identity option is associated with disciplinary group. It is worth mentioning that logistic regression is based on several tests, including Omnibus Test, Wald Test, and Nagelkerke R Square. Using these tests, logistic regression predicts the effects of various independent variables on dependent variables. It means that the logistic regression tries to test the effects of various independent variables on dependent variables by developing a hypothetical model. To this end, it develops a null hypothesis in which the independent variables are assumed to have no effect on dependent variables, which it calls it as intercept model.

2.5. Ethical Considerations & Rigor of the Study

Considering the fact that the essays analyzed did not reveal the identity information of the authors, authors' anonymity was followed. In addition, Prof. Hilary Nesi, who was the responsible for the BAWE corpus, was contacted, and she expressed her consent for using the corpus.

Confirmability and dependability of the study were ensured using a mixed-method design in which the interpretation of findings was clearly derived from the data. In other words, some part of the required data was transparently and directly tagged to various essays (e.g. age, native language, level, grade, etc.). Other extracted data including authorial identity were developed out of the essays using a transparent methodology based on which any other researcher achieves the same findings. Furthermore, to ensure the consistency of data analysis, we kept a meticulous record of the text analysis in MAXQDA2020. Credibility of the study, on the other hand, was done through analyst triangulation in which another qualified researcher reviewed the data and the developed interpretations.

3. Results

Table 3 shows the results for the intercept model, that is the Maximum Likelihood model, if only the intercept (the independent variables) is included without any of the dependent variables in the analysis.

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

Table 3. *Classification Table*[a,b]: *Data Classification Using the Null Model*

Observed			Predicted		
			identity option		Percentage Correct
			authorial	=non-authorial	
Step 0	identity option	authorial	0	237	.0
		=non-authorial	0	287	100.0
	Overall Percentage				54.8

a. Constant is included in the model.

b. The cut value is .500

Based on Table 3, 54.8% of individuals were correctly classified using the null model (everyone was classified as non-authorial) and 74.1% were correctly classified using the full model (prediction of using or non-using authorial identity) which is a large improvement as Table 4 shows.

Table 4. *Classification Table*[a]*: Data Classification Using the Full Model*

Observed			Predicted		
			identity option		Percentage Correct
			authorial	=non-authorial	
Step 1	identity option	authorial	102	135	68.3
		=non-authorial	60	227	79.2
	Overall Percentage				74,1

a. The cut value is .500

In addition to assessing the individual significance of each independent variable, the full model is tested using a Likelihood Ratio (LR) test (a test that assesses the goodness of fit of two competing statistical models based on the ratio of their likelihoods) to see if it is a significant improvement (p-value < 0.05) on the null model in the 'Model' row of the 'Omnibus Tests of Model Coefficients' table. The model was statistically significant when compared to the null model, $\chi 2\ (31) = 141.261$, $p < 0.001$ as Table 5 shows.

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

Table 5. *Omnibus Tests of Model Coefficients*

		Chi-square	df	Sig.
Step 1	Step	141.261	31	.000
	Block	141.261	31	.000
	Model	141.261	31	.000

R2 value in linear regression gives a measure of the proportion of variation of the dependent variable explained by the model. According to Table 6, the Nagelkerke R Square value is 0.469, so about 47% of the variation in identity option can be explained by the full model suggesting that predictions are fairly reliable. In other words, the considered identity variables (age, gender, native language, group, education, level, and disciplinary group) are the factors predicting authorial identity option among the students.

Table 6. *Model Summary: Predictions of Authorial Identity Option based on the Full Model*

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	854.918	.317	.469

Table 7 shows the Wald test results for each independent variable. The output presented in Table 7 shows that the p-value is lower than the cut-score. In other words, the variables "social and contextual factors", "agency", "knowledge & discourses", "gender", "group", "education", and "disciplinary group" were found to be significantly effective.

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

Table 7. *Wald test Results for Each Independent Variable*

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Social actors			3.130	2	.209			
	Social actors (1)	.123	.272	.205	1	.651	1.131	.663	1.929
	Social actors (2)	.371	.212	3.075	1	.079	1.450	.957	2.196
	Social & contextual factors			5.209	2	.044			
	Social & contextual factors (1)	.790	.349	5.121	1	.024	3.203	1.112	4.366
	Social & contextual factors (2)	.647	.357	.944	1	.031	1.415	.702	2.852
	Agency			11.491	3	.009			
	Agency (1)	.929	.718	7.226	1	.007	1.145	.036	.593
	Agency (2)	.843	.706	6.821	1	.009	3.158	.040	.631
	Agency (3)	.438	.733	11.051	1	.001	1.087	.021	.368
	Knowledge and discourses (1)	.603	.283	.133	1	.016	1.309	.637	1.930
	Age	.027	.025	1.164	1	.281	1.028	.978	1.080
	Gender (1)	.327	.213	2.353	1	.012	2.387	.913	2.106
	Group (1)	.427	.413	.303	1	.042	4.797	.355	1.789
	Education			1.483	2	.047			
	Education (1)	.419	.344	1.481	1	.024	4.520	.774	2.985

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

Education (2)	.454	.719	.125	1	.024	2.290	.315	5.282
Level			2.331	3	.507			
Level (1)	.144	.527	.074	1	.785	1.155	.411	3.247
Level (2)	.503	.558	.812	1	.368	1.654	.554	4.941
Level (3)	-.065	.824	.006	1	.937	.937	.186	4.713
Disciplinary group			8.523	3	.036			
Disciplinary group (1)	.408	.392	8.007	1	.005	1.330	.153	.711
Disciplinary group (2)	.933	.473	3.898	1	.048	2.393	.156	.993
Disciplinary group (3)	-.242	.534	5.407	1	.020	.289	.101	.823
Type of Data	.440	.591	36.581	1	.681	1.237	.363	2.345
Constant	-1.995	208.31	36.776	1	.000	.283		

These results show that social & contextual factors (p = .044), agency (p = .009), knowledge and discourses (p = .016), gender (p = .012), group (p = .042), education (p=.047), and disciplinary group (.036) added significantly to the model/prediction, but social actors (p = .209), age (p=.281), as well as type of data (p= .681) and level (p=.507) did not add significantly to the model.

The column B demonstrates the coefficient for each independent variable. Table 8 illustrates that the B coefficients for the two social & contextual factors (educational settings) and (cultural backgrounds) are significant and positive, indicating that authorial identity use is associated with educational settings and cultural backgrounds. The Exp(B), on the other hand, tells us that educational settings are more than three times (3.203) more likely than cultural backgrounds or social interactions to affect identity option among university students.

Moreover, the B coefficients for the all agencies (1-3) are significant and positive, indicating that agency is also associated with identity option. The Exp(B), on the other hand, reveals that goals and expectations (agency 2) are about three times (3.158) more likely than

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

beliefs and tenets (agency 1), and interests and preferences (agency 3) influence identity option among university students.

The variable "knowledge and discourses" was found to have a significant B coefficient which is positive, indicating that it is correlated with identity option. Meanwhile, this variable affects about 30% of identity option based on which authorial identity option is preferred by the university students in their essays. Furthermore, The B coefficients for both gender and group are significant and positive, indicating that they are associated with identity option. Likewise, females tend to use authorial identity more than two times (2.387) more likely than males. In addition, native speakers (group 1) prefer to use authorial identity about five times (4.797) more likely than their non-native speakers' counterparts. Moreover, the B coefficients for all "Education" variable are positive and significant association with identity option in which studying in UK completely or partially were found to be effective. In other words, students with the background studying solely in UK tended to use authorial identity more than four times (4.52) than their counterparts whose education is solely situated overseas than UK. Meanwhile, students who have spent part of their education in UK tended to use authorial identity more than two times (2.29) than their counterparts whose education is solely situated overseas than UK.

The B coefficients for various disciplinary groups are significant and positive, indicating that they are associated with identity option. Hence, identity option is highly related to discipline and depends on the norms and tenets of each discipline. In addition, using identity option is more than two times observed in essays on "Social Sciences" than other disciplinary groups ("Arts & Humanities" & " Life Sciences"), and on "Arts and Humanities" about 30% more likely than "Life Sciences"

In sum, a logistic regression was performed to ascertain the effects of disciplinary group", "level", "agency", "social & contextual factors", "social actors", "education", "group", "gender". "knowledge & discourses", and "type of data" on the likelihood that university students use authorial identity. The logistic regression model was statistically significant, $\chi 2(31) = 141.261$, $p < 0.001$. The model explains about 47% (Nagelkerke R2) of the variance in identity option and correctly classifies 74.1% of essays. In other words, our model can classify 74.1% of the essays in terms of identity option (authorial vs. non-authorial) by the independent variables of "disciplinary group", "level", "agency", "social & contextual factors", "social actors", "education", "group", "gender". "Knowledge & discourses", and "type of data".

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

In a further step, another binary logistic regression was performed to test if there is any relationship between identity option and the grades given to the essays. Table 8 shows the results for the intercept model.

Table 8. *Classification Table[a,b]: Data Classification Using the Null Model*

	Observed		Predicted		
			identity option		Percentage Correct
			authorial	=non-authorial	
Step 0	grade	70%-100%	0	81	.0
		=50%-69%	0	118	100.0
	Overall Percentage				59.3

a. Constant is included in the model.

b. The cut value is .500

Based on Table 8, 59.3% of individuals were correctly classified using the null model (everyone was classified as grade M), and 72% were correctly classified using the full model which is a large improvement (Table 9).

Table 9. *Classification Table[a]: Data Classification using the Full Model*

	Observed		Predicted		
			identity option		Percentage Correct
			70-100%	=50-69%	
Step 1	Grade	70-100%	272	49	84.7
		=50-69%	95	98	50.8
	Overall Percentage				72,0

a. The cut value is .500

The full model is tested using a Likelihood Ratio (LR) test to see if it is a significant improvement (p-value < 0.05) on the null model in the 'Model' row of the 'Omnibus Tests of Model Coefficients' table. The model was statistically significant when compared to the null model, $\chi 2$ (27) = 125.425, p < 0.001 as Table 10 shows.

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

Table 10. *Omnibus Tests of Model Coefficients*

		Chi-square	df	Sig.
Step 1	Step	125.425	27	.000
	Block	125.425	27	.000
	Model	125.425	27	.000

According to Table 11, the Nagelkerke R Square value is 0.362 so about 36% of the variation in the given grades can be explained by the full model suggesting that predictions are fairly reliable

Table 11. *Model Summary: Predictions of Identity Option Based on the Full Model*

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	654.318	.296	.362

Table 12 shows the Wald test results for identity option. Table 12 shows that identity option (p=012) adds significantly to the model/prediction. Table 12 also illustrates that the B coefficient for identity option is significant and positive, indicating that authorial identity use is associated with grades. The Exp(B), on the other hand, tells us that authorial is about two and a half times (2.387) more likely than non-authorial identity option to affect the grades given to the essays

Table 12. *Wald Test Results for Identity Option*

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
identity option (1)	1.440	1.91	56.581	1	.012	2.237	.161	.345
Type of data	.118	.192	.380	1	.537	1.888	.610	1.294
Constant	20.995	6027.405	.000	1	.011	1.512		

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

In short, conducting the logistic regression ascertain the effects of authorial identity on the likelihood that students are given the grade D. Likewise, the logistic regression model was statistically significant, $\chi 2(27) = 125.425$, $p < 0.001$. The model explains about 36% (Nagelkerke R2) of the variance in identity option and correctly classifies 72% of essays.

4. Discussion

In light of the results obtained, we are capable of answering the three raised questions. The first question was to determine whether there is any relationship between identity-related variables and manifestation of authorial identity in the essays written by the university students. It was found that social and contextual factors, agency, knowledge and discourses, gender, group, education, and disciplinary group added significantly to the model (prediction of using or non-using authorial identity). Meanwhile, social actors, age, as well as type of data (corpus or EFL university students' essays), and level did not add significantly to the model.

Overall, authorial identity use was found to be associated with numerous variables, namely, educational settings and cultural backgrounds (university, school, college, and cultural backgrounds), knowledge and discourses, academic/non-academic encounters, gender, group, education (studying in UKA completely or partially or overseas), and disciplinary group.

Moreover, there is a rich body of literature in which the relationships between identity and a vast verity of variables have been identified. For instance, identity was found to be associated with various individual, social, and contextual factors (Gao et al., 2007; Boonchum, 2009; Anspal et al., 2012; Park, 2012; Duff, 2013; Vasilopoulos, 2010; 2015; Wielgosz & Molyneux, 2015; Shin, 2016; Zotos, Moon, and Shultz, 2020), past experiences (Kubota, 2001; Wielgosz & Molyneux, 2015; González, Moll, & Amanti, 2005; Shin, 2012, 2016; Vasilopoulos, 2015; Wielgosz & Molyneux, 2015), and sociodemographic backgrounds (Hall & du Gay, 1996; Kibria, 2002; Pavlenko & Blackledge, 2004; Gao, 2011; Duff, 2013; Liu, 2015; Vasilopoulos, 2015; Wielgosz & Molyneux, 2015; Shin, 2016), and agency (Ahearn, 2001; Park, 2012; Duff, 2013; Vasilopoulos, 2015; Wielgosz & Molyneux, 2015). Furthermore, the results of the study did not confirm the relation between identity and relational contexts. Hence, these findings may be in contrast with the studies in which a significant relation between identity and relational contexts was found (e.g. Taylor, 2010; Vasilopoulos, 2015; Wielgosz & Molyneux, 2015; Lane et al., 2019; Martin, 2020).

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

The relationship between identity option and the grades given to the essays was also explored in order to answer the second question. Having analyzed the gathered data, the researchers found a positive and significant relationship between identity option and the grades given to the essays. Having resorted to Ivanic's (1998) arguments, we can assume that university students especially the marginalized ones (those with different sociodemographic backgrounds) have problems in portraying their self in their writing. In fact, the findings revealed that university students for whom English is considered as a native language or the ones belonging to the majority outperformed their marginalized counterparts. Such findings can be explicated in light of some studies (e.g. He, 2020; Pittam et. Al., 2009). In fact, the underrepresentation of authorial identity among non-native speakers of English or students with lower sociodemographic backgrounds is due to their unfamiliarity with using authorial identity and the obstacles they perceive in expressing their authorial identities in the required assignments (Pittam et. Al., 2009). It may also denote that the three factors discussed by Pittam et al. (2009), namely, *"confidence in writing*", *"understanding authorship*", and *"knowledge to avoid plagiarism*" are more dominant among the mainstream group (students whose native language is English and their education is completely or partially developed in the UK). He (2020) emphasizes that learners' identities especially the ones for whom English is considered as an additional language "are highly affected by their English proficiency levels, educational experience, disciplinary conventions, genre affordances, and audience awareness" (p. 506). He also adds that EFL/ESL university students "tend to present a weak authorial identity (and are) more engaged with texts than with readers and lack commitment to their claims (and their identities) are interwoven with their English proficiency levels, educational experience, disciplinary conventions, genre affordances, and audience awareness" (He, 2020, p. 506).

Finally, the study tried to see if there is any relationship between authorial identity and disciplinary group. Analyzing the data illustrated disciplinary group as a factor representing or non-representing authorial identity in the essays. In other words, the norms and tenets of each discipline affect the way authorial identity is manifested in students' writings. Likewise, as many previous studies, we can assume knowledge and discourses as a variable affecting identity in general and authorial identity in particular (Harwood, Hansen, & Lotter, 2006; Commins et al., 2015; Chesler & Young, 2007; Lotter, Harwood, & Jose, 2007; Lin, 2009; Hong & Greene, 2011; Windle & Miller, 2012; Liu, 2015; Vasilopoulos, 2015; Wielgosz & Molyneux, 2015; Shin, 2016; Meo & Tarabini, 2020; Villegas, et al., 2020; Zotos, Moon, & Shultz, 2020).

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

5. Conclusion

Authorial identity is the self-concept expressed by the author in their writings through which textual identity is developed. Likewise, authorial identity is a non-unitary notion which is multiple and rooted in various epistemological preferences based on disciplines and disciplinary groups. Consequently, we can consider subjectivity features for authorial identity, too. Furthermore, students' sociodemographic and background knowledge are two main variables affecting underrepresentation of authorial identity which may be justified through the students' unfamiliarity with representing their self or lack of confidence. Such factors may provide us with the discourse-related obstacles perceived by the students. Accordingly, authorial identity is a complex process restructured through the process of negotiation with various individual, discoursal, sociocultural, and demographic characteristics. These factors may develop individuals to consider self-legitimacy in employing authorial identity, though it is also rooted in epistemological preferences of their disciplines.

Henceforth, the findings of the study may be of interest for education in general and special education in particular. In effect, sociocultural, political, and social contexts wherein education is taking place are of great importance which affect students' identity and as a result their further academic achievements. Furthermore, learners' background knowledge including their native language, education, and disciplines are among the factors which may make differences in the way text identity is developed by students in their essays. Similarly, identity option may be one of the problems which is faced especially by overseas students. It, in turn, may affect the development of their academic discourses. Specialized pedagogies may also be suggested as an intervention to enable the students to depict their text identity. Moreover, enhancing students' skills in deploying their legitimacy in representing authorial identity may motivate them to improve their argumentative, evaluative and analyzing skills which are fundamental for academic settings. These critical skills seem to be main factors for academic achievements and performances.

References

Ahearn, L. M. (2001). Language and agency. *Annual Review of Anthropology, 30*(1), 109-137.

Anspal, T., Eisenschmidt, E., & Löfström, E. (2012). Finding myself as a teacher: Exploring the shaping of teacher identities through student teachers' narratives. *Teachers and Teaching, 18*(2), 197-216.

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

Boonchum, P. (2009). A study of self-identity changes and correlation of influential factors of Thai students studying English. *Educational Research and Review*, *4*(11), 535–548.

Bourdieu, P. (1977). The economics of linguistic exchanges. *Social Science Information,* 16, 645–668.

Chesler, M., & Young Jr, A. A. (2007). Faculty members' social identities and classroom authority. *New Directions for Teaching and Learning*, *2007*(111), 11-19.

Cummins, J. (2004). Multiliteracies pedagogy and the role of identity texts. In K. Leithwood, P. McAdie, N. Bascia, & A. Rodigue (Eds.), *Teaching for deep understanding: Towards the Ontario curriculum that we need* (pp. 68–74). Toronto: Ontario Institute for Studies in Education of the University of Toronto and the Elementary Federation of Teachers of Ontario.

Cummins, J., & Early, M. (Eds.). (2011). *Identity texts: The collaborative creation of power in multilingual schools*. Stoke-on-Trent: Trentham Books.

Cummins, J., Hu, S., Markus, P., & Kristiina Montero, M. (2015). Identity texts and academic achievement: Connecting the dots in multilingual school contexts. *TESOL Quarterly, 49*(3), 555-581.

Duff, P. A. (2013). Identity, agency, and second language acquisition. In A. Mackay & S., M., Gass (Eds.), *Handbook of second language acquisition* (pp. 410-426). London : Routledge.

Fairclough, N. (1992). Discourse and text: Linguistic and intertextual analysis within discourse analysis. *Discourse & Society*, *3*(2), 193-217.

Gao, F. (2011). Exploring the reconstruction of Chinese learners' national identities in their English-language-learning journeys in Britain. *Journal of Language, Identity & Education*, *10*(5), 287–305.

Gao, Y., Zhao, Y., Cheng, Y., & Zhou, Y. (2007). Relationship between English learning motivation types and self-identity changes among Chinese students. *TESOL Quarterly, 41*(1), 133–155.

Gee, J. P. (2004). *Situated language and learning: A critique of traditional schooling*. London: Routledge.

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

Giampapa, F. (2001). Hyphenated identities: Italian-Canadian youth and the negotiation of ethnic identities in Toronto. *International Journal of Bilingualism*, *5*(3), 279-315.

González, N., Moll, L. C., & Amanti, C. (Eds.). (2005). *Funds of knowledge: Theorizing practices in households, communities, and classrooms*. New Jersey: Lawrence Erlbaum Associates.

Hall, S., & Du Gay, P. (Eds.). (1996). *Questions of Cultural Identity*: California: SAGE Publications.

Harwood, W. S., Hansen, J., & Lotter, C. (2006). Measuring teacher beliefs about inquiry: The development of a blended qualitative/quantitative instrument. *Journal of Science Education and Technology*, *15*(1), 69-79.

Hawkins, M. R. (2005). Becoming a student: Identity work and academic literacies in early schooling. *TESOL Quarterly*, *39*(1), 59-82.

He, F. (2020). Identity Construction in Academic Writing of Student Writers Who Use English as an Additional Language: A Literature Review1. *Chinese Journal of Applied Linguistics, 43*(4), 506-524.

Hejazi, E. Lavasani, M. G. Amani, & Was, H. (2010). Academic identity status, goal orientation, and academic achievement among high school students. *Journal of Research in Education, 22*(1), 292-320.

Hong, J., & Greene, B. (2011). Hopes and fears for science teaching: The possible selves of preservice teachers in a science education program. *Journal of Science Teacher Education*, *22*(6), 491-512.

Ivanič, R. (1998). *Writing and identity* (Vol. 10). Amsterdam: John Benjamins.

Kibria, N. (2002). Of blood, belonging, and homeland trips: transnationalism and identity among second-generation Chinese and Korean Americans. In P. Levitt & M. C. Waters (Eds.), *The changing face of home: The transnational lives of the second generation* (pp. 295-311). New York: Russel Sage Foundations.

Kubota, R. (2001). Discursive construction of the images of U.S. classrooms. *TESOL Quarterly, 35*(1), 9–38.

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

Lane, A. K., Hardison, C., Simon, A., & Andrews, T. C. (2019). A model of the factors influencing teaching identity among life sciences doctoral students. *Journal of Research in Science Teaching*, *56*(2), 141-162.

Lin, L. F. (2009). Second language learners' identity toward their home culture: Adding pragmatic knowledge to language learning curriculum. *Asian Social Science*, *5*(8), 43-51.

Liu, S. (2015). Searching for a sense of place: Identity negotiation of Chinese immigrants. *International Journal of Intercultural Relations*, *46*, 26-35.

Lotter, C., Harwood, W. S., & Bonner, J. J. (2007). The influence of core teaching conceptions on teachers' use of inquiry teaching practices. *Journal of Research in Science Teaching*, *44*(9), 1318-1347.

Manyak, P. C. (2004). "What did she say?" Translation in a primary-grade English immersion class. *Multicultural Perspectives*, *6*, 12–18.

Martin, C. (2020). From LSA to teacher: the value of classroom experience in shaping a 'teacher'identity. *Support for Learning*, *35*(1), 23-42.

Meo, A., & Tarabini, A. (2020). Teachers' identities in second chance schools: A comparative analysis of Buenos Aires and Barcelona. *Teaching and Teacher Education*, *88*, 1-11.

Norton Peirce, B. (1995). Social identity, investment and language learning. *TESOL Quarterly, 29*, 9–32.

Norton, B. (1997). Language, identity, and the ownership of English. *TESOL Quarterly, 31*(3), 409-429.

Norton, B. (2013). *Identity and language learning: Extending the conversation* (2nd ed.). Bristol: Multilingual Matters.

Park, G. (2012). "I am never afraid of being recognized as an NNES": One teacher's journey in claiming and embracing her nonnative speaker identity. *TESOL Quarterly*, *46*, 127–151.

Park, H. (2012). Insight into learners' identity in the Korean English as a lingua franca context. *Journal of Language, & Education*, *11*(4), 229–246.

Pavlenko, A., & Blackledge, A. (Eds.). (2004). *Negotiation of identities in multilingual contexts* (Vol. 45). Clevedon: Multilingual Matters.

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

Pittam, G., Elander, J., Lusher, J., Fox, P., & Payne, N. (2009). Student beliefs and attitudes about authorial identity in academic writing. *Studies in Higher Education, 34*(2), 153-170.

Shin, J. (2012). Power dynamics in a multilingual Korean language classroom and pedagogical implications. In P. C. Miller, J. L. Watzke, & M. Mantero (Eds.), *Readings in language studies: Language and identity* (pp. 501–520). Grandville, MI: International Society for Language Studies.

Shin, J. (2016). Hyphenated identities of Korean heritage language learners: Marginalization, colonial discourses and internalized whiteness. *Journal of Language, Identity & Education*, *15*(1), 32–43.

Stone, L. D., & Hart, T. (2019). *Sociocultural psychology and regulatory processes in learning activity*. Cambridge: Cambridge University Press.

Toohey, K., Manyak, P., & Day, E. (2007). ESL learners in the early school years: Identity and mediated classroom practices. In J. Cummins & C. Davison (Eds.). *International handbook of English language teaching* (pp. 625–638). New York: Springer Science.

Usborne, E., & Taylor, D. M. (2010). The role of cultural identity clarity for self-concept clarity, self-esteem, and subjective well-being. *Personality and Social Psychology Bulletin*, *36*(7), 883-897.

Vasilopoulos, G. (2010). English and self-identity construction among Korean learners: A qualitative study. *외국어교육연구*, *24*(2), 101-126.

Vasilopoulos, G. (2015). Language learner investment and identity negotiation in the Korean EFL context. *Journal of Language, Identity & Education, 14*(2), 61–79.

Walker, D. (2014). *A pedagogy of powerful communication : Youth radio and radio arts in the multilingual classroom*. New York : Peter Lang.

Weedon, C. (1987). *Feminist practice and poststructuralist theory*. London: Blackwell.

Welch, O. M., & Hodges, C. R. (1997). *Standing outside on the inside: Black adolescents and the construction of academic identity*. New York: SUNY Press.

West, C. (1992, Summer). *A matter of life and death*. October, 61, 20–23.

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]

Wielgosz, M., & Molyneux, P. (2015). "You Get to Be Yourself": Visual arts programs, identity construction and learners of English as an additional language. *Journal of Language, Identity & Education, 14*(4), 275–289.

Windle, J., & Miller, J. (2012). Approaches to teaching low literacy refugee-background students. *The Australian Journal of Language and Literacy*, *35*(3), 317-333.

Winter, R. (2009). Academic manager or managed academic? Academic identity schisms in higher education. *Journal of Higher Education Policy and Management*, *31*(2), 121-131.

Zotos, E. K., Moon, A. C., & Shultz, G. V. (2020). Investigation of chemistry graduate teaching assistants' teacher knowledge and teacher identity. *Journal of Research in Science Teaching*, *57*(6), 943-967.

[Downloaded from ijal2.khu.ac.ir on 2026-11-24]