Thin-slice Judgments of English Language Teacher Success in Instruction: The Effects of Learners' Gender, Age, and Language Proficiency[1]

Nahid Soltanian[2], Ferdowsi University of Mashhad, Mashhad, Iran

Aramreza Sadeghi[3], Semnan University, Semnan, Iran

Abstract

The present study aimed to explore English as a foreign language (EFL) learners' first impressions of their language teachers. The participants were 679 male and female Iranian EFL learners from different age groups and English language proficiency levels at five language institutes and four universities. The main data collection instrument was a validated 47-item questionnaire. The findings showed that at the end of the semester, EFL learners' first impressions of their language teacher, formed based on short observations of thin slices of behavior (less than five minutes), did not change significantly. The results also indicated that there were significant differences in the extent of first impression change between male and female participants and different language proficiency levels (low- and high-proficiency). Moreover, there was no statistically significant difference between different age groups (teenagers and adults) in terms of the extent of first impression change. The pedagogical implications of the findings regarding the impacts of language learners' thin-slice judgments of the teachers' success in EFL contexts are discussed.

Resumen

El presente estudio tuvo como objetivo explorar las primeras impresiones de los estudiantes de inglés como lengua extranjera (EFL) de sus profesores de idiomas. Los participantes fueron 679 estudiantes iraníes de inglés como lengua extranjera, hombres y mujeres, de diferentes grupos de edad y niveles de dominio del idioma inglés en cinco institutos de idiomas y cuatro universidades. El principal instrumento de recolección de datos fue un cuestionario validado de 47 ítems. Los hallazgos mostraron que al final del semestre, las primeras impresiones de los estudiantes de inglés como lengua extranjera de su maestro de idiomas, formadas con base en observaciones breves (menos de cinco minutos), de comportamiento no cambiaron significativamente. Los resultados también indicaron que hubo diferencias significativas en el grado de cambio de la primera impresión entre los participantes masculinos y femeninos y los diferentes niveles de competencia lingüística (competencia baja y alta). Además, no hubo diferencias estadísticamente significativas entre los diferentes grupos de edad (adolescentes y adultos) en términos de la extensión del cambio en la primera impresión. Se discuten las implicaciones pedagógicas de los hallazgos con respecto a los impactos de los juicios finos de los estudiantes de idiomas sobre el éxito de los maestros en contextos de inglés como lengua extranjera.

Introduction

When encountering someone for the first time, human perceivers frequently employ physical attributes to form quick and automatic judgments of target persons and interpret their personality traits. These judgments are interestingly made despite the fact that each person is a source of complex and vague information and has dynamic characters (Naumann et al., 2009; Schiller et al., 2009). These rapid initial interpretations are formed through thin slices of behavior, which are brief samples of social behavior (Ambady et al., 2000). In fact, personality traits are not directly observable, but perceivers naturally use expressive behaviors such as facial expressions, gait patterns, occupational stereotypes evoked by attire, posture, speech, and patterns of touch to make inferences about other individuals (Ambady & Rosenthal, 1992; Blaskovits & Bennell, 2019; Küster, Krumhuber, & Hess, 2018).

Human perceivers' judgments and evaluations of others, made based on thin slices of even less than half a minute, are often surprisingly accurate (Rule & Ambady, 2008), persistent, and resistant to change (Schiller et al., 2009), and are not usually inferior to judgments which are made through thicker slices (Carney et al., 2007). The power of first impression (FI) must be recognized as people engage in social interactions (Wilson et al., 2011). People's FIs can profitably direct them while they navigate the social world and make sense of their environment (Alaei & Rule, 2019), and assist social perceivers in determining positive relationships and potentially damaging interactions (Ambady & Gray, 2002). They also play a significant role in individuals' feelings, attitudes, and behaviors toward others in the process of social perception (Kinley et al., 2019).

[1] This is a refereed article. Received: 27 May, 2020. Accepted: 7 September, 2020.
[2] nah.soltanian@gmail.com
[3] aramsadeghy@semnan.ac.ir

Among the wide variety of thin-slice research studies that have been conducted during the last few decades, one line has examined perceivers' judgments of teachers through thin slices of behavior (e.g., Ambady & Rosenthal, 1992, 1993). This research line indicates that what teachers do within the classroom in the initial five seconds or so can predict how they are evaluated at the end of the semester (Ambady & Rosenthal, 1992; Babad et al., 2004).

Given the practical importance of FI formation in human interactions (Kinley et al., 2019; Naumann et al., 2009) and the point that no previous thin-slice study has been conducted on English as a foreign language (EFL) learners, the present study focused on EFL learners' FIs of their language teachers. It aimed to investigate the EFL learners' judgments of their language teachers when the perceiver (the individual who makes judgements) and the target (the individual who is judged) were unacquainted and were given no considerable opportunity to interact in order to create a situation in which the judgments are formed based on the little information (e.g., physical appearance) available to perceivers. The study, thus, examined the extent to which the learners' FIs on the basis of brief observations of their teacher (thin slices) changed during the semester. Moreover, the role of the learners' age, gender, and English language proficiency level in the extent of FI change were examined. For the purposes of this study, the following research questions were addressed:

1. *To what extent do EFL learners' impressions of their language teachers differ at the beginning and at the end of the semester?*
2. *Are there any significant differences between EFL learners' degree of FI change and their demographic characteristics such as gender, age, and English proficiency level?*

Literature Review

A thin slice is described as a short and dynamic excerpt of an individual's behavior which is normally less than five minutes and is often known as a slice taken from a longer video-recording of the individual while interacting with other people or carrying out a task (Oltmanns et al., 2004). In thin-slice studies, judges, who are normally unacquainted with targets, rate the targets that have been recorded through a specific method. Targets can be recorded through different channels. Previous research has typically used nonverbal channels including the vocal (tone) and visual cues (body, face, or both of them), verbal channels including speech and transcripts, and audiovisual channels which combine the visual and verbal techniques (Ambady & Rosenthal, 1992).

Perceivers' FIs are usually examined through group consensus and overall accuracy (Krendl et al., 2014). According to Krendl et al. (2014), group consensus reflects "how strongly perceivers agree with each other in their FIs" and overall accuracy reflects "judgments that are predictive or correct at levels significantly greater than chance guessing" (p. 483). Consensus and accuracy are generally associated, but this is not always true (Rule et al., 2013), since people may be highly consensual in the judgments they make (e.g., who seems to be reliable), but these judgments might not be predictive of the target's behavior in an accurate way (e.g., who is really reliable) (Krendl et al., 2014).

Thin-slice Research in Education

Although there is a rich literature on thin-slice judgments in various domains including emotions (Ambady & Gray, 2002), sexual orientation (Rule & Ambady, 2008), visibility of social class (Bjornsdottir & Rule, 2017), individuals' antigay prejudice (Alaei & Rule, 2019), and vulnerability to victimization (Blaskovits & Bennel, 2019), few studies have focused on FI formation in education (e.g., Greenfield & Quiroz, 2013; Wentzel, 2009). The cases that follow are examples of studies which have examined thin-slicing in school and university contexts.

In a series of studies by Ambady and Rosenthal (1993) college and high school teachers were videotaped while teaching. Silent video fragments (under 30 seconds) were selected from these videotaped sessions. The instructors were rated by naïve judges on some features (e.g., being active, accepting, empathic, or competent). These ratings, which were made by judges unfamiliar with the instructors through nonverbal thin slices, significantly anticipated both overall end-of-semester student evaluations and supervisor evaluations of the teachers. In the same article, Ambady and Rosenthal (1993) reported on using thinner slices (6-second and 15-second clips) to examine if strangers' ratings of the teachers on the same dimensions could make predictions about the two above-mentioned criterion variables related to the previous studies. The results showed that the ratings had strong relationships with the criterion variables.

Babad et al. (2004) investigated the significance of students' ratings of professors' nonverbal behavior within the classroom in content-free short cases (nine seconds) in predicting end-of-semester students' ratings of teaching (SRT). They videotaped the professors in 67 courses in four situations including the first session of class, while they were giving lectures, interacting with students, and speaking about the course. Findings of the study showed that thin slices of the professors' content-free nonverbal behavior could anticipate SRT, but there were different designs for the four situations. Positive judgments of short cases of nonverbal lecturing behavior anticipated positive post-course SRT constituents connected with the professors. Moreover, positive judgments of nonverbal behavior, when the professors interacted with the students, had a negative relationship with favorable and positive SRTs. This outcome was explained by the point that SRTs had a negative relationship with the difficulty of the course, and although the professors might have done their best in their interactions with the students in difficult courses, the ratings of these courses were lower.

In Babad's (2005) study, high-school students watched ten-second silent thin slices of unfamiliar teachers' nonverbal behavior while they were giving lectures to their students. The students were asked to rate the extent to which each teacher would show equal treatment toward high- and low-achieving students in a one-on-one interaction. The students could predict which teachers had different treatment toward their low-achievement students (e.g., by having shorter duration of eye contact with them) in comparison to their high-achievement ones (e.g., by having longer duration of eye contact with them) and which teachers showed equal treatment toward their low-and high-achievement students. The ratings of these naïve judges had positive correlations with the teachers' actual bias, as shown by the evaluations of the teachers' own students.

Although there are few studies on thin slicing in education, language learning is a field which has not been dealt with in the thin-slice research and little is known about how language learners' thin-slice judgments of language teachers may change. Considering the point that FIs are significant in human interactions (Kinley et al., 2019; Naumann et al., 2009) and that a language classroom should enjoy many interactions between the teacher and the language learners (Zhao & Bitchener, 2007), the present study focuses on this context. Moreover, there are few recent thin-slice studies in the field of education. The present study addresses this gap by systematically investigating language learners' FIs of their teachers in EFL contexts.

Age, Gender, Language Proficiency and Interpretation of Expressive Behaviors

Previous studies have mentioned some potential factors which influence the interpretation of expressive behaviors in EFL classrooms including learners' age, proficiency in the target language, nativeness, and years of learning the language (Kamiya, 2019). Gender is also a factor which has been mentioned to be effective in FI formation (Carney et al., 2007; Gonçalves et al., 2014), but its effect on FI formation in the language classroom has not been investigated so far. This is the factor which is considered in this section along with the other factors.

A small number of studies have examined the impacts of age on FI formation (e.g., Franklin & Zebrowitz, 2013; Humă, 2010; Zebrowitz et al., 2013). This line of research has suggested that older and younger perceivers may think uniformly when they evaluate others' personality traits and that they are not typically different in the general consensus of the impressions, however, they might be different in the magnitude of these impressions (Petrician et al., 2014). The literature shows that the effect of age on impression formation has been corroborated by few studies (Humă, 2010; Krendl et al., 2014). On the other hand, other studies, such as Kumin and Lazar (1974) and Mohan and Helmer (1988), have shown that older perceivers are better decoders of nonverbal behaviors than younger ones. Although these studies have indicated that age is effective in decoding nonverbal behaviors, one main problem with these studies is that they have investigated the effect of age among participants whose age difference is just one year, which may have been too short to see a change. To address this inquiry, Kamiya (2018) recruited two age groups with a gap of over six years in his study. Although Kamiya (2018) showed an age effect for making more accurate judgments of the teacher, some limitations in his study make it unclear whether the adults' more accurate judgments could be ascribed to age by itself or to some other factors. Kamiya (2018) argued that the better performance of adults may have been owing to the video stimuli (short silent video clips selected from second language (L2) classrooms in a US university) with which the adults were familiar, but children were not. Moreover, since adults, in comparison to children, had higher English proficiency and had learned English for a longer period, the age effect could have been attributed to the amount of English learning or the learners' English proficiency.

Kamiya conducted another study in 2019 to examine how learner's age, English proficiency, and years of English language learning affected the accuracy of deciphering teachers' nonverbal behaviors. The participants watched silent video clips of EFL classrooms in Japanese elementary schools. They were asked to recognize if the teachers were asking a question or not. It was found that both children and adults could use the teachers' nonverbal behaviors to judge whether the teachers were asking a question greater than at chance levels. The noticeable point was that in comparison to the mean accuracies of children, those of adults were statistically higher although the clips had been selected from elementary school classrooms. The learners who had studied English for more than six years paid attention to a greater number of nonverbal behaviors to make correct judgments. This, in turn, increased the minimum accuracy of their judgments. This finding reflects a significant point in Kamiya (2018) that the higher tendency shown by adults to judge nonverbal behaviors more accurately was not due to the nature of stimuli, but to the fact that they showed more accuracy in their judgments irrespective of the kind of stimuli. Nonetheless, it is not acceptable to infer that "the older, the better" since there is a close connection between age and proficiency, years of learning, and the extent of exposure to the target language, culture, and context (Kamiya, 2019).

When social phenomena like impression formation are explored, investigating gender, as a potential influential factor in FI formation, is a common sociological practice (Humă, 2010). Although the studies on FI formation have demonstrated different results for the judgments made by different genders when they evaluate others, it has generally been indicated that female judges show more accuracy than male ones in the perception of nonverbal behaviors (Hall, 1984), especially in the case of judging emotions (Hall & Matsumoto, 2004).

In one study, Chan et al. (2011) aimed to investigate whether female perceivers, in comparison to male ones, could more accurately understand other individuals' personality, by viewing brief videos of target individuals. The findings showed that female perceivers were more accurate than male ones in normative accuracy, or their perceptions of what others were generally like. However, no significant gender differences were found among the participants with respect to distinguishing the distinct and unique personality features of other individuals or their perceptions of how others were dissimilar from the average person.

In another study, Kinley et al. (2019) investigated if perceiver gender would influence their impression of a target person wearing a hearing aid or glasses. In this study, no significant difference was found between male perceivers in their judgments of a target with or without glasses, but more differences could be identified by females between the target and control participants wearing glasses. In general, females' rating of the male target wearing glasses was more positive. Fewer gender differences were shown concerning the target with the hearing aid since female and male perceivers' ratings of this target generally concurred.

Nativeness has also been identified as influential in forming FIs. Some studies have compared gesture interpretation between natives and non-natives (those who shared or did not share common ethnic background with gesture performers). They have demonstrated that natives can interpret the meaning of gestures better than non-natives (Dahl & Ludvigsen, 2014; Mohan & Helmer, 1988) because they are more proficient than the non-natives (Kamiya, 2019). Although these studies have shown the better performance of natives in nonverbal behavior interpretation, the problem with them is that the natives enjoyed more proficiency in the language and had learned it for longer periods. Thus, it is not clear whether the better performance of native participants in gesture interpretation is due to being native, being more proficient in the target language, or being more exposed to the target language and culture over the years (Kamiya, 2019).

Kamiya (2019) hypothesized that the amount of learning English language is influential in the interpretation of nonverbal behavior among EFL learners, since in EFL contexts, the number of years (or hours) of L2 learning strongly influences the development of proficiency. However, in English as a second language (ESL) contexts, the age at which a L2 learner starts to learn an L2 is taken as a suitable predictor of his/her eventual achievement. The significance of the amount of L2 learning in language proficiency development has been shown in different studies done in EFL contexts (e.g., Ojima et al., 2011). According to Kamiya (2019), since the ability to decode nonverbal behavior improves along with the linguistic advancement and the accumulation of classroom experience, it can be argued that the

number of years spent on learning English is the most significant variable in making more accurate judgments, interpreting teacher gestures, and decoding their nonverbal behavior among EFL learners.

To add new evidence to the existing literature of thin slicing in education, the current study investigates FI formation in EFL classrooms. Previous research has considered accuracy or consensus in thin-slice judgments. This study, however, is concerned with the extent of change in language learners' FIs of language teachers' success in instruction. It focuses on language learners' judgments in face-to-face contexts among individuals at zero acquaintance. This study also represents one of the initial steps of assessing the relationship between demographic characteristics and FI formation in EFL contexts. Building on the extant literature, we speculate about the importance and contribution of several factors to thin slicing in the EFL classroom. The factors which are considered to be more relevant to this context and are addressed in this study are age, gender, and English language proficiency level. These three variables have not been fully investigated in the previous empirical thin-slice studies in EFL contexts.

Methodology

Participants

This study aimed to survey English language learners' FIs of their language teachers in an EFL context. The participants included 679 male and female EFL learners at five language institutes and four universities in Iran belonging to different language proficiency levels and age groups. The choice of the classes in these educational contexts was that the students were unfamiliar with the relevant language teacher (i.e., zero acquaintance).

Table 1 demonstrates the descriptive statistics (frequency and percentage) for the participants' age, gender, and English proficiency level. Language proficiency level was self-rated by the respondents on the basis of the tests of English language proficiency taken at their institutes or universities. To facilitate data analysis and presentation, language proficiency was reduced to two categories of high proficiency (HP) (including intermediate, high intermediate, and advanced levels) and low proficiency (LP) including basic, elementary, and pre-intermediate levels). Age was also categorized into two groups of teenagers (11–19) and adults (over 19).

Age		Gender		Language Proficiency level	
		Male	Female	LP	HP
Teenagers	200 (29.5%)	96 (14.1%)	104 (15.3%)	226 (33.2%)	45 (6.6%)
Adults	479 (70.5%)	201 (29.6%)	278 (40.9%)	161 (23.7%)	247 (36.3%)
Total	679 (100%)	297 (43.7%)	382 (56.3%)	387 (57%)	292 (43%)

Table 1: Participants' age, gender, and language proficiency Level

Table 1 indicates that the adult group represents the higher number of participation (70.5%) followed by the teenage group (29.5%) of the total participants. Furthermore, this table shows that there are more female participants than male with a ratio of 1.28 from whom 56.3% are female and 43.7% others are male. The frequency of participants from different language proficiency levels is also presented in this table. As Table 1 demonstrates, 43% and 57% of the participants belong to the HP and LP levels, respectively.

Data Collection Instrument

The present study was organized based on the model for a successful EFL teacher suggested by Moafian and Pishghadam (2009) who developed and validated the *Characteristics of Successful Teachers Questionnaire* (CSTQ). The data they gathered and analyzed produced twelve factors for the construct of teacher success. These factors included teaching accountability, interpersonal relationships, attention to all, examination, commitment, learning boosters, creating a sense of competence, teaching boosters, physical and emotional acceptance, empathy, class attendance, and dynamism.

CSTQ has been developed in the Persian language and is comprised of 47 items tapping the twelve components of language teacher success. Each item in the questionnaire is based on a five-point Likert-type scale which ranges from strongly agree to strongly disagree. The construct validity of the questionnaire has been ensured through exploratory factor analysis, and the reliability of the

questionnaire is 0.94. In the present study, Cronbach's alpha for this scale was 0.88. For participants, answering all items took about 30 minutes.

Data Collection Procedure and Data Analysis

To conduct the research, initially one of the researchers secured the permission of twelve university professors and thirteen instructors at language institutes to administer the questionnaire among the students in the early minutes of the class. This researcher was familiar with the professors and instructors whose classes had been chosen for the purpose of this study and asked them not to query about the content and purpose of the questionnaire since it was possible that this awareness would cause them not to perform authentically. The professors and instructors were also asked to leave the class at the time of the questionnaire administration.

Prior to the beginning of the class, one of the researchers talked to the language learners in the class and asked them to participate in the study. She explained that they should fill out the questionnaire based on their impressions of the teacher after they saw him/her in the early minutes (less than 5 minutes). The participants, who were free to take part based on their willingness, were informed that the teacher would leave the class before they began filling out the questionnaire.

The students responded to the CSTQ at the beginning of the first session, when they did not know the language teacher yet. After four months, at the end of the semester, the procedure was repeated; CSTQ was administered again to the same participants to evaluate the teacher at the end of the semester and make it possible to compare the language learners' initial judgments of teachers with the learners' end-of-semester ratings of the same teachers. After data collection, descriptive statistics and nonparametric tests of Wilcoxon Matched-Pairs Signed-Ranks and Mann-Whitney U in SPSS were run as the main statistical methods.

Results and Discussion

To answer the research questions of the study, it was necessary to calculate the scores from the questionnaire. In this questionnaire, the scales were arranged from 1 to 5 with strongly disagree getting 1 point and strongly agree at the other end of the scale having 5 points. Consequently, each respondent obtained a total score from the questionnaire in Time 1 (when they filled out the questionnaire for the first time in the first session, known as score 1) and Time 2 (when they filled out the same questionnaire for the second time at the end of the semester, known as score 2) (Minimum score = 47, and Maximum score = 235). In the following sections, the results for the research questions are presented and discussed. For statistical analyses, nonparametric tests were employed since the One-Sample Kolmogorov-Smirnov Test showed that the distribution of data was not normal.

Results for the First Research Question

The first research question of the study examined the extent of change in the learners' FIs of their language teacher during the semester. To examine the difference between score 1 and score 2, a nonparametric test of Wilcoxon Matched-Pairs Signed-Ranks was run.

Table 2 demonstrates the results of the test.

		N	Mean Rank	Sum of Ranks
SCORE2 - SCORE1	Negative Ranks	258[a]	380.20	98092.50
	Positive Ranks	395[b]	292.25	115438.50
	Ties	26[c]		
	Total	679		

a. SCORE2 < SCORE1

b. SCORE2 > SCORE1

c. SCORE2 = SCORE1

Table 2: Wilcoxon Matched-Pairs Signed-Ranks test for the extent of change in FI

The results in Table 2 show that the sum of the ranks for the case in which score 1 is higher than score 2 equals 98092.50 (mean rank = 380.20) and for the case in which score 2 is higher than score 1 is 115438.50 (mean rank = 292.25). In order to investigate if the difference between score 1 and score 2

was significant, the inferential section of Wilcoxon Matched-Pairs Signed-Ranks Test was run. Table 3 presents the results of this test.

	SCORE2 - SCORE1
Z	-1.801[a]
Asymp. Sig. (2-tailed)	.072

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

Table 3: Test statistics[b] for the significant difference between the two scores

According to the results shown in Table 3, since *p*-value equals 0.072 (z = -1.801, *p*> 0.05), it can be concluded that there is no significant difference between score 1 and score 2. Hence, the learners did not demonstrate a significant level of change in their FIs of the language teacher. The number of participants who exhibited a very low level of change in their FIs of the language teacher was noticeable (91.9%). This can be informative for the language teachers, especially in EFL contexts, since they can appreciate the value of the early minutes of the language teacher's entrance into the class in the first session. The findings obtained were partly compatible with those of Ambady and Rosenthal's (1993) research. In their research, the teaching performance of instructors was significantly predicted by naïve judges who viewed only 30-second (or thinner slices of 6-second and 15-second) clips of teacher nonverbal behavior. Like Ambady and Rosenthal's (1993) study, the results of the present study indicated that thin slices of behavior can convey a great wealth of information about the teachers including their instructional quality.

The results obtained in this part of our study are also consistent with those of Babad et al.'s (2004) research. In our study, thin slices of teacher behavior helped the learners to predict end-of-semester evaluations of the teacher, as Babad et al. (2004) found the prediction of SRT on the basis of the thin slices of professors' content-free nonverbal behavior in their study.

Results for the Second Research Question

The second research question of the study asked about the existence of any significant difference between the learners' degree of FI change and their demographic characteristics of gender, age, and English proficiency level. This research question consisted of three sub-questions. To facilitate data analyses and make understanding the results easier, it was divided into the three distinct null hypotheses below:

H01: There is no statistically significant difference between male and female participants and the extent they change their FIs of the language teacher.

H02: There is no statistically significant difference between the extent of Iranian EFL learners' change in FIs of the language teacher and their age.

H03: There is no statistically significant difference between the extent of Iranian EFL learners' change in FIs of the language teacher and the level of their proficiency in English.

In order to test the first null hypothesis, nonparametric test of Mann-Whitney U was used. The results are presented in Table 4.

Gender	N	Mean Rank	Sum of Ranks
Male	296	361.70	107063.00
Female	383	323.23	123797.00
Total	679		

Table 4: Mann-Whitney U Test for gender

Table 4 shows that the mean rank in the male group equals 361.70 and in the female group is 323.23. That is, the extent of FI change among males is more than that in the female group. In order to examine if this difference was significant, the inferential section of Mann-Whitney U Test was run. Table 5 presents the results of this test.

	DIF
Mann-Whitney U	50261.000
Wilcoxon W	123797.000
Z	-2.539
Asymp. Sig. (2-tailed)	.011

a. Grouping Variable: Gender

Table 5: Test statistics[a] for the significant difference between males and females

According to Table 5, the *p*-value for the difference in the extent of FI change between females and males is 0.011 (U = 50261.000, $p<0.05$). Hence, it can be said that there is a significant difference between females and males in the extent of FI change. Hence, the first null hypothesis is rejected, and it is concluded that Iranian male and female EFL learners show different levels of change in their FIs of English language teacher.

The reason Iranian male and female English language learners are statistically different in the extent of their FI change (the females showing less change in their FIs) can be that females are naturally better perceivers of individuals' personality traits since they pay more attention to external states and tend to be more accurate in their perceptions of nonverbal behaviors (Kinley et al., 2019; Murphy et al., 2003; Vogt & Colvin, 2003). Since females have the tendency to be more interpersonally oriented and typically show a higher level of interpersonal sensitivity in comparison to males, they are generally more attuned to what people are like (Moskowitz, Suh, & Desaulniers, 1994). Furthermore, biological as well as social factors may be effective in the behavioral differences between males and females (Ellis et al., 2008). Katsikitis et al. (1997) also explain that there is a gender-based pattern in the mechanisms of nonverbal communication decoding; while men employ a prototypical image to identify emotions, women typically consider different cues and analyze them one by one. Additionally, female participants may be more confident in their evaluations of others. The findings of this part can be compared with those of Chan et al. (2011) in that a gender-based pattern was also observed in the participants' evaluations of some personality traits. The results of this part of the study are not consistent with Kinley et al.'s (2019) study since in their research no significant difference was observed between males and females with regard to their competence in judging the targets based on thin slices.

To test the second null hypothesis and compare the score difference in the two age groups of teenagers and adults, the nonparametric test of Mann-Whitney U was employed. The results are demonstrated in Table 6 below.

	Age	N	Mean Rank	Sum of Ranks
DIF	Teenagers (11-19)	479	343.32	164450.50
	AdulIts (19+)	200	332.05	66409.50
	Total	679		

Table 6: Mann-Whitney U test for age

The results presented in Table 6 show that the mean rank in the teenage group equals 343.32 and in the adult group it is 332.05. So, the teenagers experienced a higher extent of FI change. However, to determine if this difference was significant, the inferential section of Mann-Whitney U Test was run. Table 7 shows the results of this test.

	DIF
Mann-Whitney U	46309.500
Wilcoxon W	66409.500
Z	-.684
Asymp. Sig. (2-tailed)	.494

a. Grouping Variable: Age

Table 7: Test Statistics[a] for the significant difference between teenagers and adults

As Table 7 demonstrates, the *p*-value for the score difference between the teenage group and the adult group is 0.494 (U = 46309.5, p>0.05). Therefore, although the adult group exhibited less change in the FI of the language teacher, it can be said that there was no significant difference in the extent of FI change between the participants of these two age groups. This finding indicates that the second null hypothesis is supported.

The reason older and younger judges did not show any significant difference in the extent of their FI change and were similar in their perceptions may be that every age has its own advantages to infer the personality traits. Although adult learners are more cognitively mature in comparison to teenagers, teenagers also pay attention to subtle cues to decode others' expressive behaviours. The ability to decode expressive behaviours initiates from early childhood and individuals at all age groups use this capacity to infer others' personality traits. Apparently, they employ the same cues of physical appearance to form judgments of others, at least in zero acquaintance contexts (Petrician et al., 2014). As they state, null age impacts have been shown concerning the use of facial appearance cues or subtle emotional cues by younger and older perceivers to decode expressive behavior since they use these cues similarly to infer personality.

The results obtained in this section of the study are in line with many prior studies which investigated the effect of age on FI formation and found that younger and older perceivers may act rather similarly when they evaluate individuals' personality traits (e.g., Humă, 2010; Langlois et al., 2000). Although research into the potential effects of age on inferring personality traits is rather scarce, the available studies show that personality judgments, both in reference to the self and other persons, are not mainly affected by the effects associated with age (Keightley et al., 2006; Ruby et al., 2009; Zebrowitz et al., 2013). Even in some cases it has been shown that older adults are seemingly more lenient judges in comparison to their younger counterparts, both of their own and of close others' personality (Grady et al., 2012), and strangers' personality traits (Zebrowitz et al., 2013).

Similar to the current research, Humă's (2010) study did not indicate the significant effect of the observer's age on the impression formation of the targets. Moreover, Kamiya's (2019) research showed that although the mean accuracies of adult participants were statistically higher than those of the children in judging teacher nonverbal behavior, both children and adults could judge the teacher's nonverbal behavior better than at chance levels.

To test the third null hypothesis, the extent of FI change was compared between the LP and HP groups of learners. To examine and compare the score difference in these two groups, the nonparametric test of Mann-Whitney U was used. Table 8 shows the results of this test.

	Language Proficiency	N	Mean Rank	Sum of Ranks
DIF	LP	387	354.48	137182.00
	HP	292	320.82	93678.00
	Total	679		

Table 8: Mann-Whitney U test for language proficiency

The results in Table 8 indicate that the mean rank in the LP group equals 354.48 and in the HP group it is 320.82. In other words, the extent of FI change in the LP group is more than that of the HP group. To see if the difference between the LP and HP groups was significant, the inferential section of Mann-Whitney U Test was considered. Table 9 demonstrates the results of this test.

	DIF
Mann-Whitney U	50900.000
Wilcoxon W	93678.000
Z	-2.218
Asymp. Sig. (2-tailed)	.027

a. Grouping Variable: LP

Table 9: Test statistics[a] for the significant difference between LP and HP groups

According to the results shown in Table 9, it can be said that the p-value for the difference in the extent of FI change between the LP and HP groups is 0.027 ($U = 50900$, $p<0.05$). So, it can be claimed that there is a significant difference between the LP and HP groups in their extents of FI change. Based on this finding, the third null hypothesis of the study is rejected.

The reason HP EFL students showed less change in their FIs appears to be that the HP group had spent more time in EFL classes, so they enjoyed more linguistic advancement, further exposure to English language, and more classroom experiences of learning English. The longer years of learning English in EFL contexts, and proficiency development and linguistic advancement in turn (Muñoz, 2006; Ojima et al., 2011), contributes to helping EFL learners to possess appropriate knowledge of cues and improve the ability of correctly interpreting teacher gestures (Kamiya, 2019; Walsh, 2011). So, these EFL learners can pay attention to a larger number of nonverbal behaviors to interpret the teacher's nonverbal behavior (Kamiya, 2019) and form FIs based on their exposure to thin slices shown by the teacher. Partly consistent with the findings of Kamiya's (2019) study, the results of the present study indicated that the HP learners exhibited less change in their FIs of the language teacher. In spite of the fact that Kamiya's (2019) research approach was somewhat different from our research– he concentrated on the accuracy of the interpretation of chiefly nonverbal behaviors among EFL learners–his work was very helpful.

Conclusions and Implications

Following the assumption that FIs often help people to form accurate impressions not only about social information such as personality and emotions, but also about the individuals' effectiveness and success, this study aimed to survey FI formation in an EFL context. To this end, we compared the impressions formed of language teachers at the beginning and the end of the semester. We also examined the effects of the three demographic characteristics including gender, age, and language proficiency on FI formation.

The results of the current quantitative-based survey indicated that the majority of Iranian EFL learners in the study did not change their FIs of the language teacher during the semester. Moreover, the difference between the extent of FI change of male and female participants was significant with males showing more change in their FIs up to the end of the semester. The same significant difference was also observed between the LP and HP levels of language proficiency with the first group demonstrating more change in their FIs of the language teacher. The results also indicated that age was not a crucial factor in the extent of FI change among the subjects.

The results obtained in this study have significant implications for education in general and language education in EFL contexts in particular, since English language course and language teachers are typically considered to be different from other courses and teachers of other subjects (Borg, 2006; Brosh, 1996). A pedagogical implication of the findings is that some stable qualities of the instructor are unconsciously conveyed in the first session. Students form rapid impressions of their teacher, whether there is verbal or nonverbal interaction between them. They employ numerous cues (subtle thin slices of behavior reflected by the teacher) to recognize the teacher's personality traits and his/her success in instruction in the first-class session of the course and do not easily change their FIs. This indicates the potential effect of the first day of class on students' impressions of a course and the significance of how teachers can manipulate the social environment in a way to help students benefit by enhanced motivation, improved attitudes, and increased learning (Wilson et al., 2011). Considering that teachers' expressive cues may be used for or against them, what and how a teacher communicates in the early minutes of the first session can affect the students' FIs positively or negatively, and this can subsequently affect the climate of the classroom, the students' behavior and learning outcomes, and the teacher's instructional effectiveness (Haleta, 1996).

A further implication of this study is that although little evidence is available to show that the improvement of teachers' effectiveness in teaching can occur by being trained in nonverbal skills (Ambady & Rosenthal, 1993), teachers can at least be trained to send proper cues to their students in the first session and do not cause fixed inaccurate FIs, and inaccurate end-of-semester evaluations in turn, by showing expressive cues which may be the result of sheer boredom, anger, anxiety, or other affective factors which are seen just at special moments.

In closing, the present researchers used a validated questionnaire in their research. Although utilizing questionnaires is considered the standard approach to personality assessment and has been accepted as a valid and reliable self-report method (Vazire, 2006), more qualitative studies are required to expand the

scope of research in this area. In future studies on FI formation, quantitative approaches can be accompanied by qualitative ones, such as interviews, to obtain richer results in this regard. The reasons these Iranian EFL learners did not demonstrate a significant level of change in FI can be various and more in-depth mixed-methods studies are required to specifically investigate the associated reasons. Now that the general image of the participants' FIs of their language teachers is clear, a narrower and deeper study embracing interviews with language learners can provide a more complete source of data. The urgency of further studies is needed to examine how paying attention to language learners' FIs of the teacher can affect the process of language learning. Further research is warranted to examine the effects of students' FIs of the teacher on the teaching process of teachers in the future. Future researchers can do this through investigating teachers' attitudes. If teachers are informed of the results of their students' FIs, there will be a better chance of changing inappropriate behaviors into learning-conducive opportunities. The present study used the 5-min observations of language teachers in a real situation as thin slices for the formation of FI. Future studies can employ other channels of behavior such as a combination of teaching clips and real observations of the teacher to investigate whether these results are corroborated. Moreover, as one main benefit of electronic communication is decreasing potential negative stereotypes of ethnicity or culture (Wilson et al., 2011), future researchers can consider the impacts of electronic communication between teacher and students on FI formation.

References

Alaei, R., & Rule, N. O. (2019). People can accurately (but not adaptively) judge strangers' antigay prejudice from faces. *Journal of Nonverbal Behavior*, *43*(3), 397–409. https://doi.org/10.1007/s10919-019-00305-2

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In M. P. Zanna (Ed.), *Advances in experimental social psychology (*Vol. 32, pp. 201–271). Academic Press.

Ambady, N., & Gray, H. M. (2002). On being sad and mistaken: Mood effects on the accuracy of thin-slice judgments. *Journal of Personality and Social Psychology, 83*(4), 947–961. https://doi.org/10.1037//0022-3514.83.4.947

Ambady, N., & Rosenthal, R. (1992). Thin slices of behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*(2), 256–274. https://doi.org/10.1037/0033-2909.111.2.256

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology, 64*(3), 431–441. https://doi.org/10.1037/0022-3514.64.3.431

Babad, E. (2005). Guessing teachers' differential treatment of high- and low-achievers from thin slices of their public lecturing behavior. *Journal of Nonverbal Behavior, 29*(2), 125–134. https://doi.org/10.1007/s10919-005-2744-y

Babad, E., Avni-Babad, D., & Rosenthal, R. (2004). Prediction of students' evaluations from brief instances of professors' nonverbal behavior in defined instructional situations. *Social Psychology of Education, 7*, 3–33. https://doi.org/10.1023/B:SPOE.0000010672.97522.c5

Bjornsdottir, R. T., & Rule, N. O. (2017). The visibility of social class from facial cues. *Journal of Personality and Social Psychology*, *113*(4), 530–546. https://doi.org/10.1037/pspa0000091

Blaskovits, B., & Bennell, C. (2019). Are we revealing hidden aspects of our personality when we walk? *Journal of Nonverbal Behavior*, *43*(3), 329–356. https://doi.org/10.1007/s10919-019-00302-5

Borg, S. (2006). The distinctive characteristics of foreign language teachers. *Language Teaching Research, 10*(1), 3-31. https://doi.org/10.1191%2F1362168806lr182oa

Brosh, H. (1996). Perceived characteristics of an effective language teacher. *Foreign Language Annals, 29*(2), 25-38. https://doi.org/10.1111/j.1944-9720.1996.tb02322.x

Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality, 41*(5), 1054–1072. https://doi.org/10.1016/j.jrp.2007.01.004

Chan, M., Rogers, K. H., Parisotto, K. L., & Biesanz, J. C. (2011). Forming first impressions: The role of gender and normative accuracy in personality perception. *Journal of Research in Personality, 45*(1), 117–120. https://doi.org/10.1016/j.jrp.2010.11.001

Dahl, T. I., & Ludvigsen, S. (2014). How I see what you're saying: The role of gestures in native and foreign language listening comprehension. *The Modern Language Journal, 98*(3), 813–833. https://doi.org/10.1111/j.1540-4781.2014.12124.x

Ellis, L., Hershberger, S., Field, E., Wersinger, S., Pellis, S., Greary, D., Palmer, C., Hoyenga, K., Hetsroni, A. & Karadi, K. (2008). *Sex differences: Summarizing more than a century of scientific research*. Psychology Press.

Franklin, R. G., Jr., & Zebrowitz, L. A. (2013). Older adults' trait impressions of faces are sensitive to subtle resemblance to emotions. *Journal of Nonverbal Behavior, 37*(3), 139–151. https://doi.org/10.1007/s10919-013-0150-4

Gonçalves, G., Gomes, A., Ferrão, M. C., Parreira, T., dos Santos, J. V., Giger, J-C., & Martins, A. T. (2014). Once upon a face: The effect of eye size, observer and stimulus gender on impression formation. *Current Psychology*, *34*(1), 112–120. https://doi.org/10.1007/s12144-014-9244-3

Grady, C. L., Grigg, O., & Ng, C. (2012). Age differences in default and reward networks during processing of personally relevant information. *Neuropsychologia, 50*(7), 1682–1697. https://doi.org/10.1016/j.neuropsychologia.2012.03.024

Greenfield, P. M., & Quiroz, B. (2013). Context and culture in the socialization and development of personal achievement values: Comparing Latino immigrant families, European American families, and elementary school teachers. *Journal of Applied Developmental Psychology, 34(2),* 108 –118. https://doi.org/10.1016/j.appdev.2012.11.002

Haleta, L. L. (1996). Student perceptions of teachers' use of language: The effects of powerful and powerless language on impression formation and uncertainty. *Communication Education, 45*(1), 16–28. https://doi.org/10.1080/03634529609379029

Hall, J. (1984). *Nonverbal sex differences: Communication accuracy and expressive style*. Johns Hopkins University Press.

Hall, J. A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion, 4*(2), 201–206. https://doi.org/10.1037/1528-3542.4.2.201

Humă, B. (2010). Gender differences in impression formation. *Journal of Comparative Research in Anthropology and Society, 1*(1), 57–72. http://doctorat.sas.unibuc.ro/wp-content/uploads/2010/10/BogdanaHuma_Compaso2010.pdf

Kamiya, N. (2018). The effect of learner age on the interpretation of the nonverbal behaviors of teachers and other students in identifying questions in the L2 classroom. *Language Teaching Research, 22*(1), 47–64. https://doi.org/10.1177%2F1362168816658303

Kamiya, N. (2019). What factors affect learners' ability to interpret nonverbal behaviors in EFL classrooms? *Journal of Nonverbal Behavior*, *43*(3), 283–307. https://doi.org/10.1007/s10919-019-00297-z

Katsikitis, M., Pilowsky, I., & Innes, J. M. (1997). Encoding and decoding of facial expression. *Journal of General Psychology*, *124*(4), 357–370. https://doi.org/10.1080/00221309709595565

Keightley, M. L., Winocur, G., Burianova, H., Hongwanishkul, D., & Grady, C. L. (2006). Age effects on social cognition: Faces tell a different story. *Psychology and Aging, 21*(3), 558–572. https://doi.org/10.1037/0882-7974.21.3.558

Kinley, T., Strübel, J., & Amlani, A. (2019). Impression formation of male and female millennial students wearing eyeglasses or hearing aids. *Journal of Nonverbal Behavior*, *43*, 357–379. https://doi.org/10.1007/s10919-019-00296-0

Krendl, A. C., Rule, N. O., & Ambady, N. (2014). Does aging impair first impression accuracy? Differentiating emotion recognition from complex social inferences. *Psychology and Aging, 29*(3), 482–490. https://psycnet.apa.org/doi/10.1037/a0037146

Kumin, L., & Lazar, M. (1974). Gestural communication in preschool children. *Perceptual and Motor Skills, 38*(3), 708–710. https://doi.org/10.2466/pms.1974.38.3.708

Küster, D., Krumhuber, E. G., & Hess, U. (2018). You are what you wear: Unless you moved—Effects of attire and posture on person perception. *Journal of Nonverbal Behavior*, *43*, 23–38. https://doi.org/10.1007/s10919-018-0286-3

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin, 126*, 390–423. https://doi.org/10.1037/0033-2909.126.3.390

Moafian, F., & Pishghadam, R. (2009). Construct validation of a questionnaire on characteristics of successful EFL teachers. *Pazhuhesh-e Zabanhe-ye Khareji Journal*, *54*, 127–142.

Mohan, B., & Helmer, S. (1988). Context and second language development: Preschoolers' comprehension of gestures. *Applied Linguistics, 9*(3), 275–292. https://doi.org/10.1093/applin/9.3.275

Moskowitz, D. S., Suh, E. J., & Desaulniers, J. (1994). Situational influences on gender differences in agency and communion. *Journal of Personality and Social Psychology, 66*(4), 753–761. https://doi.org/10.1037//0022-3514.66.4.753

Muñoz, C. (Ed.). (2006). *Age and rate of foreign language learning*. Multilingual Matters.

Murphy, N. A., Hall, J. A., & Colvin, C. R. (2003). Accurate intelligence assessments in social interactions: Mediators of gender effects. *Journal of Personality*, *71*(3), 465–493. https://doi.org/10.1111/1467-6494.7103008

Naumann, L.P., Vazire, S., Rentfrow, P. J., & Gosling, S. D. (2009). Personality judgments based on physical appearance. *Personality and Social Psychology Bulletin, 35*(12), 1661–1671. https://doi.org/10.1177/0146167209346309

Ojima, S., Matsuba-Kurita, H., Nakamura, N., Hoshino, T., & Hagiwara, H. (2011). Age and amount of exposure to a foreign language during childhood: Behavioral and ERP data on the semantic comprehension of spoken English by Japanese children. *Neuroscience Research, 70*(2), 197–205. https://doi.org/10.1016/j.neures.2011.01.018

Oltmanns, T. F., Friedman, J. N. W., Fiedler, E. R., & Turkheimer, E. (2004). Perceptions of people with personality disorders based on thin slices of behavior. *Journal of Research in Personality, 38*(3), 216–229. https://doi.org/10.1016/S0092-6566(03)00066-7

Petrician, R., Todorov, A., & Grady, C. (2014). Personality at face value: Facial appearance predicts self and other personality judgments among strangers and spouses. *Journal of Nonverbal Behavior, 38*(2), 259–277. https://doi.org/10.1007/s10919-014-0175-3

Ruby, P., Collette, F., D'Argembeau, A., Peters, F., Degueldre, C., Balteau, E., Luxen, A., Maquet, P., & Salmon, E. (2009). Perspective taking to assess self-personality: What's modified in Alzheimer's disease? *Neurobiology of Aging, 30*(10), 1637–1651. https://doi.org/10.1016/j.neurobiolaging.2007.12.014

Rule, N. O., & Ambady, N. (2008). Brief exposures: Male sexual orientation is accurately perceived at 50 ms. *Journal of Experimental Social Psychology*, *44*(4), 1100–1105. https://doi.org/10.1016/j.jesp.2007.12.001

Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. *Journal of Personality and Social Psychology, 104*(3), 409-426. https://doi.org/10.1037/a0031050

Schiller, D., Freeman, J. B., Mitchell, J. P., Uleman, J. S., & Phelps, E. A. (2009). A neural mechanism of first impressions. *Nature Neuroscience, 12*(4), 508–514. https://doi.org/10.1038/nn.2278

Vazire, S. (2006). Informant reports: A cheap, fast, easy method for personality assessment. *Journal of Research in Personality, 40*(5), 472–481. https://doi.org/10.1016/j.jrp.2005.03.003

Vogt, D. S., & Colvin, C. R. (2003). Interpersonal orientation and the accuracy of personality judgments. *Journal of Personality*, *71*(2), 267-295. https://doi.org/10.1111/1467-6494.7102005

Walsh, S. (2011). *Exploring classroom discourse: Language in action*. Routledge.

Wentzel, K. R. (2009). Students' relationships with teachers as motivational contexts. In K. R. Wenzel & D. B. Miele (Eds.), *Handbook of motivation at school* (pp. 301–322). Routledge/Taylor & Francis.

Wilson, J. H., Naufel, K. Z., & Hackey, A. A. (2011). A social look at student-instructor interactions. In D. Mashek, & E. Y. Hammer (Eds.), *Empirical research in teaching and learning: Contributions from social psychology* (pp. 32-50). Willey-Blackwell.

Zebrowitz, L. A., Franklin, R. G., Jr., Hillman, S., & Boc, H. (2013). Older and younger adults' first impressions from faces: Similar in agreement but different in positivity. *Psychology and Aging, 28*(1), 202–212. https://doi.org/10.1037/a0030927

Zhao, S. Y., & Bitchener, J. (2007). Incidental focus on form in teacher–learner and learner–learner interactions. *System*, *35*(4), 431–447. https://doi.org/10.1016/j.system.2007.04.004