

RESEARCH

Open Access



Development and validation of an English teachers' visual literacy scale for smartphone photography grounded in social semiotic theory

Ayad Kamalvand¹ and Reza Khany^{1*}

*Correspondence:
r.khany@ilam.ac.ir

¹ Department of English
Language and Literature, Ilam
University, Pajooheh Street Ilam,
Ilam, Iran

Abstract

Smartphones and the literacy to harness the educational affordances of photographs currently rest on the bedrock of second language (L2) education. Building on social semiotic theory, this study developed and validated a visual literacy scale for smartphone photography (VLS4SP). Despite the importance of visuals and smartphones, no valid scale is available in the domain of English language teaching (ELT) to measure L2 teachers' VL. In developing and validating the new VLS4SP, initially, some items were generated and tested for validity. Second, the scale was tested through rigorous psychometric analyses. The results of structural equation modeling (SEM) suggested the existence of 28 items that represent the established conceptualization of VL in the literature and the three areas of meaning metafunctions in the social semiotic theory: representational, interactive, and compositional. Overall, the study tapped upon two crucial technological and literacy factors in ELT which paves the way for further research and the applications of the scale in ELT.

Keywords: VLS4SP, Visual literacy, Social semiotics theory, Smartphone photography, ELT

Background

L2 instructors and learners while relying on sensory organs — eyes for visual reception, ears for auditory perception, and tongue for verbal expression — which are instrumental in linguistic interactions use photographs as semiotic resources for meaning making. Besides, they extensively use smartphones either for regular communication or visualization of their daily and educational activities. Today, among many other uses, instructors and learners leverage their social media accounts as bloggers or Instagram influencers for teaching and learning purposes. Content that teachers and learners produce and consume is mainly facilitated by smartphone cameras. The affordances inherent to smartphone camera systems have broadened the scope of semiotic resources accessible to individuals, facilitating the visualization of their existence through the production of photographic still images, the subsequent

sharing of these visual artifacts within online environments, the acquisition of external feedback, and the opportunity for reflective engagement with these representations at later junctures.

Digital photography as the common practice of visually oriented learners (Brumberger, 2011) plays a canonical role in actualizing VL in ELT. Gnach et al., (2022) assert that visuals are the privileged mode of communication in ELT and indispensable parts of teaching and learning (Newman & Ogle, 2019). Visuals contribute to L2 pedagogy by aiding learners to employ a broader language repertoire in their oral narratives (Lee, 2024). Additionally, they generate interest and motivation, provide context, and enhance comprehension when combined with verbal instruction. Research has underscored the value of visuals in language teaching and learning, with learners exhibiting more meaningful learning when exposed to a combination of visual and verbal instructional modalities (Mayer, 2002), in contrast to scenarios devoid of such visual aids, which may impair learners' comprehension of the subject matter. In light of the aforementioned transformations, VL has emerged as a field of inquiry dedicated to equipping contemporary students, often referred to as "viewer learners," with the skills necessary to engage meaningfully with their surroundings (Romero & Bobkina, 2017). Despite the well-established educational value of visual resources, research indicates that educational systems have not fully acknowledged the importance of VL (Avgerinou & Ericson, 1997), and teachers have yet to fully exploit their potential Donaghy and Xerri (2017). Specifically, L2 instructors have not adequately harnessed the power of images to inspire discussion, creativity, and new ideas. Moreover, disposition in ELT still reckon written and spoken modes sufficient to communicate and teach; however, this dogma has been questioned (Katsampoxaki-Hodgetts, 2024), and it has been emphasized that "it is time to unsettle this commonsense notion" (Kress, 2000). Although VL is the most essential literacy for the twenty-first century (Matusiak, 2020) and for ELT practice, little is written about how to evaluate L2 instructors' VL smartphone photography. With a visual social semiotic frame of reference and established principles of scale development, the present study attempts to develop and validate a scale for evaluating L2 teachers' VL for smartphone photography. The development of this scale facilitates the advancement of ELT pedagogy by enabling the diagnosis of instructors' VL and guiding their efforts to take informed steps toward supporting learners in acquiring knowledge in novel and more efficacious ways through engagement with photographic media (Ng, 2006). Once measured, such a tool may help understand how L2 instructors interact with visual content and the multimodality of teaching and learning materials.

Smartphone photography adds another meaning-making resource to the L2 instructors' communicative repertoire (Hall, 2018). The affordance of smartphones is particularly significant in the context of L2 acquisition, which has evolved in tandem with rapid technological advancements. Language learning entails mastery of semantics and syntax but mastery of other semiotic resources (Hall, 2018; Kárpáti & Schönau, 2022). To be a successful instructor in this digitally/visually mediated context requires literacy which enables educators to interact effectively with visual resources. Recognizing this, UNESCO (2023) and the European Network for Visual Literacy (ENViL) have emphasized the role of VL in education. Common European Framework of References for Visual Literacy (CEFR-VL) has been one of ENViL's initiatives to break down VL into its

constitutes (Schönau et al., 2020). Through these efforts, ENViL underscores the importance of visual literacy in our contemporary society.

A key argument developed in this study is the necessity for VL to work with smartphone photography for L2 instructors. VL not only augments the instructors' understanding and working with photographs but also changes the understanding of the activity of L2 teaching (Hall, 2018). Given that teaching is a multimodal endeavor (Hall, 2018; Lim et al., 2022), teaching L2 learners within a pedagogical context marked by widespread smartphone photography necessitates a focus on VL and a reevaluation of L2 instructors' instructional approaches (Miller, 2014; Thomson & Uddin, 2023). To address the expanded notion of literacy, this study developed and validated a VLS4SP grounded in social semiotic theory (Kress & Van Leeuwen, 2021). The impetus for the current study arose from the long-standing need within ELT to establish robust, theoretically grounded, and empirically validated instruments for the multimodal assessment of instructional practices (Lim & Kessler, 2024). The significance of this research lies in its capacity to enhance the field's evaluation methods, enabling assessors to develop more sophisticated insights into instructors' diverse pedagogical practices and, subsequently, to offer more tailored and effective feedback that fosters professional growth and development.

Theoretical grounding: social semiotics

According to social semiotics, visual resources function as semiotic signs that convey a range of meanings. The theory sees the relationship between the signifier and the signified motivated. Sign makers intentionally select available resources to construct their messages, rather than being constrained by any particular system of communication. The defining characteristic that distinguishes social semiotic theory from other forms of semiotics is its emphasis on sign-making rather than sign use (Kress, 2009). This field of study believes that semiotic resources serve as choice options within systems to make meaning (Dressen-Hammouda & Wigham, 2022; Kress & Van Leeuwen, 2021). Signs are continually newly made and arise from the interests of the sign-makers (Kress, 2009; Van Leeuwen, 2005). From a methodological standpoint, social semiotics provides a set of analytical tools to deconstruct visual significations according to three types of metafunctions: representational, interactive, and compositional (Kress & Van Leeuwen, 2021). The theory can significantly contribute to VL by equipping L2 instructors with tripartite-metafunctional systems of meaning to effectuate the five generic sub-competencies in the sub-domain of "producing" which are the competency to present one's own images and the competency to evaluate one's own images and image-making processes, as well as some principles for reading photographs (Barrett, 2010) such as "all images require interpretation," "photographs carry more credibility and require interpretation," and "feelings are guides to interpretations." Social semiotics recognizes that all communication, including visual communication, involves interpretation. Photographs are not self-explanatory; they require active the ability the deconstruct them into their elements. The theory views photographs as tools used by sign-makers to serve social needs. It extends L2 instructors' act of photography beyond mere observation, equipping them to discern credibility, intent, and underlying messages in photographs. Social semiotics recognizes photographs as a key mode of communication. It encourages sign-makers

to express their interpretations through language, enhancing L2 instructors' interactions with visuals. In social semiotics, realizations of meaning occur through three systems of representational, interactive, and compositional metafunctions. Each of these overarching metafunctions is composed of constituent subsystems, which serve as the semiotic resources upon which sign-makers draw to visualize and communicate their intended meanings with specific effects.

Representational metafunction refers to a system of visual choices that sign-makers deploy to depict participants (depicted people, places, and things), process (the represented actions of the visualized participants), and circumstances (the place where these actions occur). It focuses on *what*: what represented participants are represented through different visual resources (Höglund, 2022). Representational metafunction is realized through narrative and conceptual structures. Narrative structures depict participants in action-reaction processes. The main element of representational meaning is vectors which connect the represented participants. On the other hand, the conceptual structure lacks vectors; thus, no dynamicity is seen in a visual layout. Conceptual representations represent participants through classification, analytical, and symbolic structures.

Interactive meanings concern the interpersonal relations among represented participants and the viewers and how the relations are constructed (Höglund, 2022). It is realized through systems of contact, social distance, power relations, involvement, and attitude. Contact is constructed by the direct and indirect gazes; social distance by shot distance (close, medium, and long shots); power relation by vertical camera angle (high or low angles); involvement by horizontal camera angle (frontal or oblique angles); and attitude by the choice of perspective. Another aspect of interactive meaning is the validity or credibility of visual layouts. Credibility can be judged by validity markers, including color, details, contextualization, illumination, brightness, and depth.

The compositional meaning of an image integrates the representational and interactive meanings of an image to construct a meaningful whole. Composition exploits three interrelated systems to relate representational and interactive meanings: (a) information value, (b) framing, and (c) salience. Information value concerns the place of participants in a visual configuration. The placement of resources on the left and right creates the sense of new and given information and on top and at the bottom creates the information value of ideal and real. Framing isolates or connects represented participants through using white space between applying color contrast. Salience is a measure of the degree to which an element draws attention to itself. It is realized through size, sharpness of focus, tonal and color contrast, perspective, and placement in the visual field.

Defining VL

VL is an essential component of multiliteracy in the twenty-first century (The New London Group, 2000). VL is often described as an overarching term that encompasses the knowledge and skills needed to effectively use and understand visual information (Avgerinou & Pettersson, 2011; Brumberger, 2019; Elkins, 2009). This knowledge includes both an interpretative and a productive component. Although the ability to analyze and interpret visuals is critical, it is not by itself sufficient for full VL; it must be accompanied by some ability to create visual material (Brumberger, 2011). For

Felten (2008), living in a visual world, however, does not mean instructors and students naturally possess sophisticated VL skills. Rather, it involves the ability to understand, produce, and use culturally significant images, objects, and visible actions. For some, VL is defined as the ability to comprehend, generate, and utilize images to communicate, think, and learn through visual means (Stokes, 2002). The construct for other researchers is coupled with heightened conscious awareness, that is VL is the ability to analyze visuals in terms of their historical context, how they are produced, their impact on society, and the moral responsibilities of their producers (Curtis, 1987; Messaris & Moriarty, 2005). Contemporary scholars view VL as a set of abilities that are largely acquired, namely, the ability to understand (read) and use (write) images, as well as the ability to reason and learn in terms of images (Avgerinou, 2007). Yenawine (1997) associates meaning finding with VL. For him, VL involves a set of skills ranging from identification of the image to more sophisticated interpretation that acts upon contextual, metaphoric, and philosophical levels.

Within the educational context, Ausburn and Ausburn (1978) state that VL can be defined as a body of enabling skills that help individuals perceive and use visuals for purposeful interpersonal communications. Further, they point out that visuals are a language with its syntax, grammar, and vocabulary by which a person can interpret and compose visual messages. Likewise, Gates (2004) speaks of VL as a language with a syntax to create educationally functioning pictures, and their meanings need to be negotiated by learners and instructors. Goin (2001) asserts that learners require the understanding that visual language employs syntactic rules. Moline (2023) argues that VL is more than just symbol recognition. It is the skill to interpret and share the meanings of images, leading to the development of “intelligent vision” (p. 7). The Association of College and Research Libraries (ACRL) has defined VL as “a set of abilities that enables an individual to effectively find, interpret, evaluate, use, and create images and visual media” (Hattwig et al., 2013).

Social semiotics suggests that images are constructed in a way comparable to language and can also be deconstructed into the choices sign-makers make to configure meaning. Hence, the ability to read and write visuals is a crucial prerequisite for social participation in an era loaded with visuals (Hermans & Schönau, 2022). VL concerns the ability of a sign-maker to manipulate signs to create new meanings (Bopry, 1994). This ability lets the sign-maker be open to novel discoveries (Bopry, 1994). Kress and Van Leeuwen (2021) classify VL into old and new (p. 21) where the old VL refers to the era when visuals played a subservient role in communication and the new is the modern era in which visuals have gained an important role. Generally, in Kress and Van Leeuwen’s grammar of visuals, the ability to produce photographs would fall under the category of design, and the ability to interpret photos would fall under the category of interpretation. Design includes the choices made in creating a photograph, such as representational, interactive, and compositional meanings. Interpretation refers to the understanding and analysis of the visual, including the decoding of its meaning and the context in which it was created.

Overall, VL is an important skill for L2 educators. This skill requires the educators’ proficiency to create, interpret, and evaluate visuals to participate fully in a *bain*

d'images (Avgerinou, 2009), or image bath, and it is increasingly recognized as a key component of education.

Research questions

- RQ 1: What are the key dimensions that constitute the construct of VLS4SP?
- RQ 2: What are the psychometric properties of the scale developed to measure English teachers' VL for smartphone photography?

Methodology

Study design

The study adopted the established psychometric principles suggested by Hinkin (2005) and DeVellis and Thorpe (2021) to develop the scale. The process involved scale development and scale validation. The scale development phase included the defining VL, a pilot study, and the content validity of the developed. The scale validation phase included exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and model fit using multiple fit indices, including χ^2 test, Tucker-Lewis index (TLI), Bentler's Comparative Fit Index (CFI), Hoelter's "critical N," root-mean-square error of approximation (RMSEA), and goodness-of-fit index (GFI).

Phase 1

Initially, based on the conceptualization of the VL informed by social semiotic theory, an initial pool of items was generated. Since the literature on VL has already provided a theoretical definition of VL and since social semiotics has delineated the components of VL, the study adopted a deductive approach to generate the items (Hinkin, 2005). Hence, to ensure that the items on the scale were representative of the construct and grounded in the theory, the relevant literature was extensively reviewed, and then a theoretical definition of the construct was formulated. Secondly, item wording was conducted. This step covered issues related to the simplicity and consistency of the items (Hinkin, 2005). The scale also was checked for ambiguity and double-barreled items that could lead to the respondents' confusion. Thirdly, the number of items was determined. This step was guided by the principle of retaining a pool of items that exceeds the target length of the final scale by approximately 50% (DeVellis & Thorpe, 2021) to achieve a higher internal. For item scaling, the participants were asked to answer the items using a 5-point Likert-style rating system. Typically, 5 to 7 points are used to increase the reliability of the Likert scales, according to common practices (DeVellis & Thorpe, 2021; Hinkin, 2005). Hence, the items were evaluated on 5-point Likert-type scales ranging from *strongly disagree* to *strongly agree*. After the participants have provided their responses using the 5-point Likert-style rating system, each response is assigned a score from 1 (strongly disagree) to 5 (strongly agree). The total score for each participant is then calculated by summing the scores of all the items. In the case of missing responses, the total score is calculated based on the items that were answered. This scoring protocol ensures that the responses are interpreted consistently across all participants, enhancing the reliability of the scale. The concluding stage of the "Phase 1" entailed the assessment of face and

content validity to ascertain the alignment of the items with its purported measurement objectives. At this stage, the questionnaire was administered to the first group of participants ($n = 30$). The participants were asked to review the items and provide feedback on whether they understood the statements questions and if any statements were unclear. Then, a panelist of experts in social semiotics and applied linguistics checked the items. The experts were chosen based on their extensive knowledge and experience in the fields of social semiotics and applied linguistics. This included the founder of the theory used in this study, who brought invaluable insights and depth to the process. This stage aimed to solicit feedback as well as to check for domain definition, domain representation, and domain relevance (Sireci, 1998).

Phase 2

In this step, the study utilized SEM using the software package AMOS to assess the reliability and validity of the VLS4SP. First, the internal consistency reliability of the scale by calculating Cronbach's alpha coefficient was done. Then, EFA was performed initially to explore the underlying factor structure of the scale. After identifying the factor structure through EFA, a CFA was performed to confirm the factor structure of the scale. Model fit was evaluated using multiple fit indices, including χ^2 test, TLI, Bentler's CFI, Hoelter's "critical N," RMSEA, and GFI.

Results

Demographic features

Pilot phase

The total number of Iranian English instructors is kept confidential by Iranian general education and not publicly available. However, efforts were made to ensure that the sample is diverse and representative of various regions and teaching contexts within Iran. To be eligible for this study, participants needed to have 5 years of experience teaching English. The pilot testing sample for this study consisted of 30 Iranian English instructors. The educational qualifications of the participants were as follows: 4 participants had PhD degrees, 15 participants had MA degrees, and 11 participants had BA degrees. The gender breakdown of the pilot sample was 45% male (13 participants) and 55% female (17 participants). The age range of the pilot phase participants was 25 to 55 years old. All participants in the pilot phase had a minimum of 5 years of English teaching experience. The pilot phase participants were recruited from various regions within Iran to ensure a diverse and representative sample. The gender ratio was 45% male and 55% female. The pilot testing evaluated the scale's clarity, specificity of directions, and preliminary psychometric properties. Further, the test-retest reliability of the scale was assessed using the pilot group of participants. The same test was administered to these participants at two different points in time. The correlation between the two sets of scores was calculated to assess the reliability of the test over time. The results showed a high correlation between the scores from the two administrations ($r = 0.85$, $p < 0.01$), indicating strong test-retest reliability. This suggests that the scale provides consistent results over time when the construct being measured remains stable.

Main phase

The main study's sample consisted of 274 English instructors with 30 holding PhDs, 106 holding BAs, and 138 holding MAs. The gender ratio was 42.7% male and 57.3% female. The main phase participants were individuals between the ages of 25 and 55. The main phase sample was also drawn from different regions across Iran to maintain a diverse representation. The scale's format was organized into two discrete parts: a demographic data section and a 5-point Likert scale section for assessing participant responses. Participants were informed of the confidentiality and research purposes of their responses. The scale was distributed through two methods: in-person or electronically using Google Forms.

Reliability assessment

A reliability assessment was conducted to evaluate the internal consistency of the scale. Cronbach's alpha coefficient was computed for the remaining 65 questions, which measures the extent to which all items in the questionnaire measure the same construct or dimension. The obtained coefficient of 0.974 indicated good internal reliability and consistency of the items, surpassing the recommended threshold of 0.65 (DeVellis & Thorpe, 2021).

Content validity assessment

Initially, 71 items were generated according to reviewing the literature. The items then were presented to the expert panel of the study to assess the content and establish its validity. After gathering feedback and rating data from the experts, the content validity index for each item was calculated. Upon analyzing the expert ratings, it was found that six items were deemed unsuitable for the study and were eliminated from the initial pool. Furthermore, based on the experts' recommendations, some of the items were modified and rephrased to enhance their clarity and specificity.

EFA

Prior to performing EFA, the Kaiser–Meyer–Olkin (KMO) for measuring the adequacy of the sampling and Bartlett's test of sphericity for factorability of the correlation matrix were assessed. The KMO measure was found to be 0.89 which indicates the sampling was adequate (Shrestha, 2021). Bartlett's test of sphericity was statistically significant ($\chi^2 = 18339.799$, $df = 2485$, $p < 0.000$) indicating highly acceptable factorability of the data and correlation matrix (Shrestha, 2021).

The study used both the scree plot and the Kaiser-Guttman method to determine the optimal number of factors to retain. The scree plot (Fig. 1) shows a distinct break after the third factor, indicating the retention of three factors. Then, the Kaiser-Guttman criterion, which suggested retaining factors with eigenvalues greater than 2.5, was used as the corresponding criterion value. Based on this criterion, the first three factors had eigenvalues greater than 2.5, while the fourth factor had an eigenvalue of 2.189, which is less than 2.5. Therefore, the first three factors were retained. As shown in Table 1, the three factors accounted for a total of 48.48% of the variance in the data. Factor 1 explained 38.54% of the variance, factor 2 explained 5.68% of the variance,

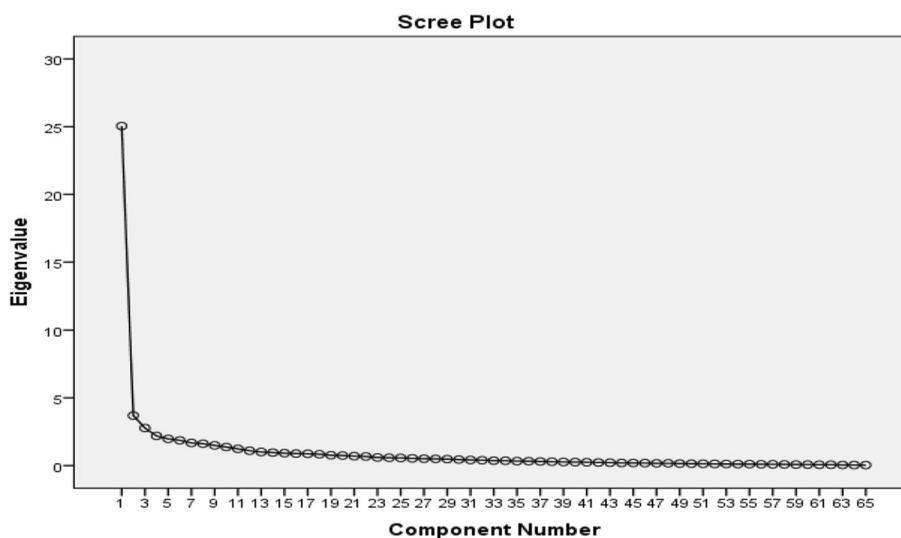


Fig. 1 Scree plot

Table 1 Total variance explained

Component	Initial eigenvalues	Extraction sums of squared loadings	Rotation sums of squared loadings	% of variance explained	Cumulative % explained
1	25.04	25.04	13.00	38.53	38.53
2	3.69	3.69	11.13	5.67	44.21
3	2.77	2.77	7.37	4.26	48.48

and factor 3 explained 4.27% of the variance. The study also conducted a rotation of the factor structure, which resulted in the extraction of the same three factors. The rotated solution accounted for a total of 48.48% of the variance in the data. Factor 1 explained 20.01% of the variance, factor 2 explained 17.13% of the variance, and factor 3 explained 11.34% of the variance. After factor rotation, items loaded on the three factors with loadings less than 0.5 were deleted.

A matrix of factor loadings (Table 2) was generated from the initial factor analysis before rotation. The matrix shows the factor loadings of the items in the questionnaire across the three components extracted. The items with higher factor loadings were considered to be more strongly related to the underlying factor.

Next, to improve the interpretability of the factor structure after the initial factor extraction, factor analysis was performed using principal component analysis with varimax rotation and Kaiser normalization. The rotation was conducted with eight iterations until convergence was achieved. The rotated component matrix was generated to improve the interpretability of the factor structure. Table 3 shows the rotated component matrix, which displays the factor loadings for each variable on each component after rotation. Variables with loadings greater than 0.5 were considered to have high associations with a particular component. The first component had high loadings from items 12, 14, 19, 22, 26, 27, 32, 42, 51, and 63, which were interpreted as representing *events and concepts*. The second component had high loadings from

Table 2 Component matrix

Item	Component		
	1	2	3
2	0.66	0.16	0.07
3	0.48	0.24	0.26
4	0.34	0.49	0.18
5	0.45	0.40	0.09
6	0.47	0.44	-0.18
7	0.54	-0.29	-0.01
8	0.47	0.14	0.20
9	0.32	0.58	-0.06
10	0.36	0.34	0.12
11	0.48	0.34	-0.32
12	0.74	-0.18	0.27
13	0.68	-0.01	-0.38
14	0.70	-0.27	0.25
15	0.62	0.31	0.25
16	0.62	0.05	-0.09
17	0.74	0.14	0.039
18	0.71	-0.03	-0.25
19	0.62	-0.14	0.34
20	0.69	0.22	-0.04
21	0.62	-0.11	0.12
22	0.60	-0.02	0.38
23	0.68	0.05	-0.03
24	0.58	0.16	0.24
25	0.65	0.26	0.02
26	0.51	-0.115	0.49
27	0.72	0.02	0.00
28	0.67	-0.09	0.07
32	0.64	-0.08	0.31
34	0.66	-0.23	-0.30
35	0.64	-0.09	-0.03
36	0.64	-0.28	-0.05
37	0.71	-0.01	0.08
38	0.62	-0.26	-0.17
39	0.46	0.42	-0.34
40	0.62	0.00	-0.27
41	0.65	0.00	0.11
42	0.67	-0.203	0.23
43	0.73	-0.22	-0.18
44	0.67	-0.14	-0.38
45	0.69	0.22	-0.27
46	0.71	-0.16	0.06
49	0.63	-0.24	-0.33
47	0.69	-0.09	-0.07
48	0.64	-0.00	0.19
50	0.71	-0.05	-0.13
51	0.65	-0.10	0.02
52	0.50	-0.33	-0.15
54	0.16	0.62	-0.09

Table 2 (continued)

Item	Component		
	1	2	3
55	0.68	−0.06	0.13
56	0.54	0.51	−0.07
57	0.68	0.08	−0.28
58	0.65	0.02	0.21
59	0.58	−0.13	−0.24
60	0.64	−0.25	−0.24
61	0.73	0.10	0.00
62	0.59	0.11	−0.15
63	0.61	−0.04	0.19
64	0.50	−0.11	−0.08
65	0.64	−0.15	0.17
66	0.64	0.11	0.20
67	0.58	−0.22	−0.15
68	0.68	−0.20	−0.02
69	0.62	0.04	0.19
70	0.64	0.26	−0.05
71	0.59	−0.29	−0.04

items 13, 18, 34, 40, 44, 46, 49, 50, 57, 62, and 67, which were interpreted as representing *relation with the viewer*. The third component had moderate loadings from items 15, 17, 20, 24, 37, 39, and 56 representing *syntax of space*. The rotated component matrix was used to identify and interpret the underlying constructs or dimensions represented by the factors. The 28 items that construct the VLS4SP are given in Appendix Table 11.

With the data in Table 3, the name of each factor was assigned (Table 4).

A factor correlation test was conducted to assess the relationships between the factors identified in the factor analysis. Table 5 displays the correlation matrix for the three factors. The results indicate that factor 1 was significantly positively correlated with factor 2 ($r=0.733$, $p<0.01$) and factor 3 ($r=0.757$, $p<0.01$). Factor 2 was also significantly positively correlated with factor 3 ($r=0.727$, $p<0.01$). These findings suggest that there are meaningful relationships between the factors and support our interpretation of the underlying constructs represented by the factors.

Indices of the first factor Table 6 shows the results of a one-sample t -test conducted on Factor 1. The test was conducted to determine whether the mean scores of the items in Factor 1 significantly differ from the hypothetical value of 3, responses indicating positive strength of agreement. The results show that the mean scores for all items in Factor 1 were significantly higher than the hypothetical value of 3, with p -values less than 0.001 and mean differences ranging from 0.547 to 0.785. These findings suggest that respondents on average had higher scores on Factor 1 compared to the hypothetical value. The confidence intervals for the mean differences ranged from 0.43 to 0.89, indicating that the differences are likely to be statistically significant.

Table 3 Rotated component matrix

Item	Component		
	1	2	3
12		0.69	
14		0.66	
19		0.66	
22		0.67	
26		0.71	
27		0.43	
32		0.65	
42		0.62	
51		0.43	
63		0.53	
15			0.49
17			0.43
20			0.49
24			0.35
37			0.26
39			0.62
56			0.69
13	0.69		
18	0.64		
34	0.73		
40	0.58		
44	0.74		
46	0.50		
49	0.72		
50	0.58		
57	0.59		
62	0.43		
68	0.56		

Table 4 Name of each factor and the related items

Name of the factor	Factor	Number of the items	Items
Relation with the viewer	1	11	13, 18, 34, 40, 44, 46, 49, 50, 57, 62, 67
Events and concepts	2	10	12, 14, 19, 22, 26, 27, 32, 42, 51, 63
Syntax of space	3	7	15, 17, 20, 24, 37, 39, 56

Table 5 Factor correlation test

		Factor 1	Factor 2	Factor 3
Factor 1	Pearson correlation	1		
	Sig. (2-tailed)			
Factor 2	Pearson correlation	0.733 ^a	1	
	Sig. (2-tailed)	0.000		
Factor 3	Pearson correlation	0.757 ^a	0.727 ^a	1
	Sig. (2-tailed)	0.000	0.000	

^a Correlation is significant at the 0.01 level (2-tailed)

Table 6 The descriptive statistics for factor 1

Item	Mean	Test value = 3					
		<i>T</i>	<i>df</i>	Sig. (2-tailed)	Mean difference	95% confidence interval of the difference	
						Lower	Upper
13	3.63	11.199	273	0.000	0.631	0.52	0.74
18	3.78	17.277	273	0.000	0.785	0.70	0.87
34	3.56	10.622	273	0.000	0.562	0.46	0.67
40	3.55	9.397	273	0.000	0.547	0.43	0.66
44	3.65	13.029	273	0.000	0.653	0.55	0.75
46	3.78	15.089	273	0.000	0.785	0.68	0.89
49	3.64	12.566	273	0.000	0.639	0.54	0.74
50	3.61	10.559	273	0.000	0.609	0.50	0.72
57	3.65	12.795	273	0.000	0.646	0.55	0.75
62	3.66	12.347	273	0.000	0.657	0.55	0.76
67	3.76	14.170	273	0.000	0.755	0.65	0.86

Table 7 The descriptive statistics for factor 2

Item	Mean	Test value = 3					
		<i>T</i>	<i>df</i>	Sig. (2-tailed)	Mean difference	95% confidence interval of the difference	
						Lower	Upper
12	3.70	12.283	273	0.000	0.701	0.59	0.81
14	3.81	14.935	273	0.000	0.810	0.70	0.92
19	3.86	16.545	273	.0000	0.861	0.76	0.96
22	3.67	11.444	273	0.000	0.668	0.55	0.78
26	3.69	11.171	273	0.000	0.693	0.57	0.82
27	3.73	14.407	273	0.000	0.730	0.63	0.83
32	3.76	15.819	273	0.000	0.755	0.66	0.85
42	3.83	15.751	273	0.000	0.828	0.72	0.93
51	3.75	14.572	273	0.000	0.752	0.65	0.85
63	3.78	14.266	273	0.000	0.785	0.68	0.89
63	3.78	14.266	273	0.000	0.785	0.68	0.89

Indices of the second factor Table 7 presents the results of a one-sample *t*-test for Factor 2. The mean score for each item is shown, along with the test value (set at 3), the *t*-value, degrees of freedom (*df*), and the significance level. The mean difference and 95% confidence interval for the difference are also provided. All items had significantly higher mean scores than the test value of 3, with mean differences ranging from 0.668 to 0.861. These results suggest that participants rated Factor 2 items positively, with mean scores ranging from 3.67 to 3.86.

Table 8 The descriptive statistics for factor 3

Item	Mean	Test value = 3					
		<i>T</i>	<i>df</i>	Sig. (2-tailed)	Mean difference	95% confidence interval of the difference	
						Lower	Upper
15	3.80	16.377	273	0.000	0.796	0.70	0.89
17	3.64	13.349	273	0.000	0.639	0.54	0.73
20	3.71	14.263	273	0.000	0.712	0.61	0.81
24	3.85	16.093	273	0.000	0.850	0.75	0.95
37	3.65	11.936	273	0.000	0.653	0.55	0.76
39	4.13	26.605	273	0.000	1.131	1.05	1.22
56	3.76	14.966	273	0.000	0.755	0.66	0.85

Indices of the third factor Table 8 presents the results of the one sample *t*-test for Factor 3. The mean, *t*-value, *df*, significance level, mean difference, and 95% confidence interval of the difference are reported for each item. The test value was set at 3. All items showed a significant mean difference from the test value of 3 ($p < 0.05$). The mean differences ranged from 0.639 to 1.131. The 95% confidence intervals of the difference ranged from 0.54 to 1.22.

CFA

To validate the factor structure identified through EFA, a CFA was conducted using AMOS. The structural model posited three latent constructs as factors and items as observed variables depicted in Fig. 2. The fit of the model was evaluated using a suite of indices: the χ^2 test, NFI, CFI, Hoelter's critical N, RMSEA, and GFI. The χ^2 test yielded a value of 0.314, indicating a good fit at the significance level of $\alpha = 0.05$. This value suggests that the model's predictions are consistent with the observed data. The NFI and CFI values were both above 0.90, reflecting a satisfactory fit of the model. Hoelter's critical N was calculated to be 200, confirming that the sample size is adequate for the model. The RMSEA was 0.049, which is within the acceptable range, indicating a good fit. Lastly, the GFI stood at 0.950, demonstrating the model's ability to reproduce the observed covariance matrix effectively. The model demonstrated a good fit to the data, as indicated by the fit indices meeting or exceeding their respective acceptance thresholds (Table 9). This supports the factor structure identified through EFA, suggesting that the model is an appropriate representation of the data.

In general, it was found that the path (Table 10) from VL to relation with the viewer at the 5% level was positive and significant, and the path from VL to events and concepts at the 5% level was also positive and significant, and the path from VL to the syntax of space at the 5% level was positive and significant. Additionally, the factors of relation with the viewer, events and concepts, and syntax of space had a significant and mutual effect on each other. Overall, the proposed model, which is the result of factor analysis, was confirmed.

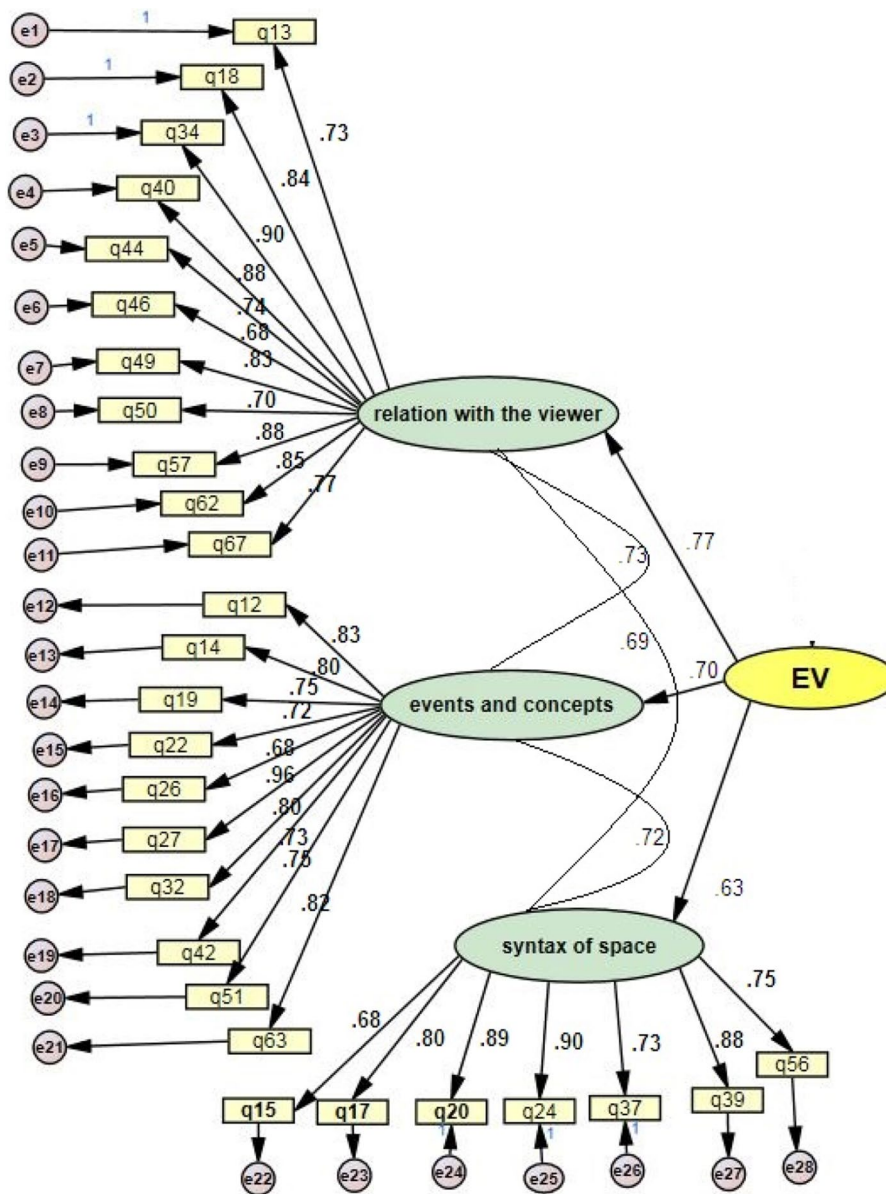


Fig. 2 Structural model of VLS4SP

Table 9 Evaluation of model fit

Criteria	Level	Interpretation
χ^2 test	0.314	Excellent fit at a significance level of $\alpha = 0.001$
NFI	0.930	Excellent fit (criteria greater than 0.90)
CFI	0.910	Excellent fit (criteria greater than 0.90)
Hoelter's "critical N"	0.840	Excellent fit (criteria greater than 0.70)
RMSEA	0.049	Excellent fit (criteria less than or equal to 0.05)
GFI	0.950	Excellent fit (criteria greater than 0.90)

Table 10 Indices for the path

Path	Estimate	p-value	Result
VL → relation with the viewer	0.77	***	Significant
VL → visualizing events and concepts	0.70	***	Significant
VL → syntax of space	0.63	***	Significant
Relation with the viewer ↔ visualizing events and concepts	0.73	***	Significant
Syntax of space ↔ visualizing events and concepts	0.72	***	Significant
Syntax of space ↔ relation with the viewer	0.69	***	Significant

***indicates a p-value less than 0.05, signifying a statistically significant result for the path estimates

Discussion

The present study attempted to develop and validate an English instructors' VL scale for smartphone photography by implying the tenets of social semiotic theory. The results of the study demonstrate that the developed scale has satisfactory validity and reliability, indicating that it is an effective tool for assessing English instructors' VLS4SP. By drawing on this theory, the study was able to develop a scale that reflects the key elements of the production and interpretation of photographs and provides a standardized framework for assessing instructors' VLS4SP. Consistent with the levels of meanings in the social semiotic theory, the results of the EFA provided support for the proposed three-factor model, which was further tested through a CFA using AMOS. The results of the CFA demonstrated an excellent fit of the proposed model to the data, as evidenced by the high values obtained for each of the fit indices. The scale comprises three factors: (a) relation with the viewer, (b) visualizing events and concepts, and (c) syntax of space with 28 items (Appendix Table 11). The first factor, in relation with the viewer, corresponds to interactive metafunction. It entails the knowledge of engaging viewers with the represented participants through the motivated use of the camera lenses to enact systems of contact, social distance, and attitude (Kress & Van Leeuwen, 2021). This factor serves to assess educators' familiarity with encoding demand or offer images. Further, it aids in deducing from taken photographs the educators' understanding of systems of the size of the frame to visualize intimacy, individuality, and personal and impersonal relationships (Jewitt & Oyama, 2001). Furthermore, the discernment of subjectivity or objectivity, as reflected in participants' detachment, power, equality, action, and knowledge orientations through camera angle, constitutes additional aspects assessable by this element. The second factor, events and concepts, aligns with the representational meaning in social semiotic theory and indicates the knowledge to use smartphone cameras to visually construct narrative and conceptual representations. It evaluates the sign-makers' ability to capture the essence of events, participants, and environment in the image. This factor taps the educators' awareness to illustrate the subject or object doing/receiving activities in a particular setting. It also serves to evaluate instructors' ability to visualize the participant as a "more or less stable and timeless essence" within the systems of classification, analytic, or symbolic structures (Kress & Van Leeuwen, 2021). The third

factor, the syntax of space which correlates with compositional meaning, refers to the knowledge of compositional principles or placing the represented participants in different spaces of a photograph. This factor facilitates evaluating educators' cognizance of information value, salience, and framing. The positioning of participants within the visual frame (information value), the extent to which they dominate the composition (salience), and the degree of their integration with additional components within the visual space (framing) collectively constitute the compositional metafunction.

Developments in technology suggest that the new VL represents a departure from the old VL and has implications for how instructors understand visual communication (Kress & Van Leeuwen, 2021). In the new VL, visuals are no longer the unstructured replication of reality, secondary to the construction of meaning (Kress, 2009). VL calls for educators who are aware of the affordances of visuals and are able to take and interpret photographs carrying intended meanings. In education, this requires a shift in teaching practices and pedagogy, as educators need to help students navigate the complex and evolving landscape of visual communication (Winstanley et al., 2024). In this way, instructors can help students develop critical thinking and analytical skills, as well as an appreciation for the power and potential of visual language. Today, English in the digitally mediated world is learned and taught through a multiplicity of modes that extend beyond modes of speech and writing (Tudini & Liddicoat, 2024). Language instructors work in a time when in 1 h more photographs are produced than were produced in the entire nineteenth century (Cubitt & Cartwright, 2018). Further, smartphones have diversified L2 learning materials and promoted learners to the producers of learning content (Huynh et al., 2022). Consequently, a tectonic shift (Kress, 1999) has happened in educational content from the medium of book to the medium of screen (Kress, 2003), demanding instructors to recognize VL as essential literacy and education to revise pedagogy (Domínguez Romero & Bobkina, 2021).

Integrating mobile devices into education has sparked heated debates among researchers. While some argue that mobile devices can enhance language skills (Xueting Ye & Shi, 2023; Yeşilel, 2022), others contend that their use can disturb the "fragile ecology" (Merchant, 2012) of classrooms. Similarly, despite the surfeit of visuals, integration of VL into ELT can be challenging. However, incorporating VLS4SP into language teaching presents both challenges and opportunities for educators. VL has been largely ignored in ELT (Donaghy & Xerri, 2017). However, the presented VLS4SP offers a promising tool for assessing instructors' abilities and raising their conscience about using photographs in ELT. The scale can aid instructors in mindfully integrating photographs into L2 education. Moreover, it helps instructors to have the needed theoretically driven vision (Ng, 2006) by which they prioritize the use of theoretically captured photographs that facilitate and intensify language learning (Wasilewska, 2017). Social semiotics theory can enrich instructors' vision for using the lenses of smartphones for pedagogical purposes. It promotes instructors' understanding that photographs are not passive reflections of reality but, rather, constructed artifacts that reflect the cultural and social practices of sign-makers. As such, instructors can use smartphone photography to encourage

students to critically examine and interpret visual images in order to understand the social and cultural meanings they convey.

The three factors of the VLS4SP can serve as a framework for guiding instructors in selecting and creating photographs that align with their instructional objectives and the learners' cultural backgrounds. The interplay between language learning and culture has been widely addressed in second-language acquisition (SLA). Culture is a reservoir of semiotic resources of a particular group of people. Individuals use the resources according to their own interests to settle on the intended meaning (Kress & Van Leeuwen, 2021). The three factors of the VLS4SP provide ELT with a systematic approach to turning photographs into semiotic modes (Danielsson & Selander, 2021). The scale raises instructors' awareness of their own culturally bound resources which in turn might lead to cultivating cross-cultural competence in L2 learners (Stefanenko & Kupavskaya, 2012). This, in turn, can lead to a more inclusive and culturally responsive approach to language teaching and learning. L2 instructors' and learners' culture has marginal status in SLA research. The VLS4SP can set the stage for future inquiries into the impact of semiotic resources.

Conclusion

In conclusion, this study has developed and validated a scale for evaluating English language instructors' VL for smartphone photography utilizing a social semiotic theoretical framework. The findings of this investigation attest to the reliability and validity of the scale as a measure of VL for smartphone photography. The instrument may serve as a benchmark for gauging VL for smartphone photography among English educators across diverse geographic and cultural contexts. The outcomes of this research may be employed as a point of reference for future inquiries seeking to assess VL among disparate populations of instructors and learners. Given the ubiquity and potency of visual media, facilitated by the proliferation of mobile devices and hosting platforms, ELT educators are compelled to cultivate their VL to effectively adapt and integrate photographic perspectives into a range of language learning tasks. The attainment of VL represents an imperative for instructors striving to remain current with and effectively engage learners' burgeoning interest and expertise in utilizing visual resources.

Limitations of the study

The VLS4SP in this study was developed within a single context. Other researchers could repeat this study in other contexts. New studies in new contexts can use this scale to refine the measure. The VLS4SP developed in this study is beginning to understand the English instructors' VL for smartphone photography. Future studies could expand the scope by including L2 instructors from various regions around the world. This would not only diversify the sample but also enhance the generalizability of the scale, making it more universally applicable and robust across different cultural and educational contexts. This broader approach could provide more comprehensive insights into VL in ELT.

Appendix

Table 11 Items constructing the VLS4SP

Items representing factor 1

- 13. I know how to take/interpret photos that visualize subjective/objective relations between the photographed people/objects and the viewer
- 18. I know how to take/interpret photos that visualize an imaginary relation between the photographed people/objects and the viewer
- 34. I know how to take/interpret photos that directly/indirectly address the viewer
- 40. I know how to take/interpret photos that use different camera angles intended to affect attitudes between the photographed people/objects and the viewer
- 46. I know how to take/interpret photos that photograph people/objects that seem to offer information to the viewer
- 49. I know how to take/interpret photos that visualize the concept of power between the photographed people/objects and the viewers
- 50. I know how to take/interpret photos that accurately visualize the subject matter and evoke a realistic feeling
- 57. I know how to take/interpret photos that visualize people/objects as parts of a whole class
- 62. I know how to take/interpret photos that use close/medium/long shots to visualize the social distance between the photographed people/objects and the viewer
- 67. I know how to take/interpret photos that intend to visualize people/objects as symbols

Items representing factor 2

- 12. I know how to take/interpret photos of people/objects that visualize a social class membership
- 14. I know how to take/interpret photos that visualize people/objects associated with a special idea or concept
- 19. I know how to take/interpret photos that deliberately do not include certain people/objects
- 22. I know how to take/interpret photos that try to visualize who or what the photographed people/objects are
- 26. I know how to take/interpret photos that visualize the process of social relations between photographed people/objects
- 27. I know how to take/interpret photos that visualize people/objects with some features and exclude some other features
- 32. I know how to take/interpret photos that lay claim upon the viewers asking the viewers to reflect on what the photographed people/objects want
- 42. I know how to take/interpret photos that convey the roles of photographed people/objects in an activity
- 51. I know how to take/interpret photos that visualize people/objects as typical of a particular group
- 63. I know how to take/interpret photos that intend to narrate an action taking place between the photographed people/objects

Items representing factor 3

- 15. I know how to take/interpret photos that place photographed people/objects in different spaces of photos to achieve an intended meaningful whole
- 17. I know how to take/interpret photos that visualize the relative significance of represented participants, with the most important elements depicted prominently and the least important elements relegated to a secondary or tertiary role
- 20. I know how to take/interpret photos that visualize photographed people/objects in layouts creating senses of connection or disconnection
- 24. I know how to take/interpret photos that intend to visualize photographed people/objects as different units of meaning
- 37. I know how to take/interpret photos that make the photographed people/objects more noticeable among other elements
- 39. I know how to take/interpret photos that visualize photographed people/objects as new/given bits of information
- 56. I know how to take/interpret photos that visualize the photographed people/objects to stand out as separate units or a group of the related unit

Abbreviations

VL	Visual literacy
VLS4SP	Visual literacy scale for smartphone photography
ELT	English language teaching
SEM	Structural equation modeling
L2	Second language
ACRL	Association of College and Research Libraries
EFA	Exploratory factor analysis
CFA	Confirmatory factor analysis
TLI	Tucker-Lewis index
CFI	Comparative fit index

RMSEA	Root-mean-square error of approximation
GFI	Goodness-of-fit index
KMO	Kaiser-Meyer-Olkin
SLA	Second-language acquisition

Acknowledgements

We would like to acknowledge the significant contributions made by Professor Theo Van Leeuwen to this study. We would also like to thank all language teachers who participated in this study. We would also like to acknowledge the editor and the reviewers of this journal for encouraging our work to be much better.

Authors' contributions

The authors confirm contribution to the paper as follows: study conception and design, AK; data collection, AK; analysis and interpretation of results, both authors; and drafting and revising the manuscript, RK. Both authors reviewed the results and approved the final version of the manuscript.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Availability of data and materials

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 12 April 2024 Accepted: 26 July 2024

Published online: 20 August 2024

References

- Ausburn, L. J., & Ausburn, F. B. (1978). Visual literacy: Background, theory and practice. *Programmed Learning and Educational Technology*, 15(4), 291–297. <https://doi.org/10.1080/0033039780150405>
- Avgerinou, M. D. (2007). Towards a visual literacy index. *Journal of Visual Literacy*, 27(1), 29–46. <https://doi.org/10.1080/23796529.2007.11674644>
- Avgerinou, M. D. (2009). Re-viewing visual literacy in the “bain d’images” era. *TechTrends*, 53(2), 28–34. <https://doi.org/10.1007/s11528-009-0264-z>
- Avgerinou, M., & Ericson, J. (1997). A review of the concept of visual literacy. *British Journal of Educational Technology*, 28(4), 280–291. <https://doi.org/10.1111/1467-8535.00035>
- Avgerinou, M. D., & Pettersson, R. (2011). Toward a cohesive theory of visual literacy. *Journal of Visual Literacy*, 30(2), 1–19. <https://doi.org/10.1080/23796529.2011.11674687>
- Barrett, T. (2010). Principles for interpreting photographs. In J. Swinnen & L. Deneulin (Eds.), *The weight of photography* (pp. 147–172). Belgium
- Bopry, J. (1994). Visual literacy in education: A semiotic perspective. *Journal of Visual Literacy*, 14(1), 35–49. <https://doi.org/10.1080/23796529.1994.11674488>
- Brumberger, E. (2011). Visual literacy and the digital native: An examination of the millennial learner. *Journal of Visual Literacy*, 30(1), 19–47. <https://doi.org/10.1080/23796529.2011.11674683>
- Brumberger, E. (2019). Past, present, future: Mapping the research in visual literacy. *Journal of Visual Literacy*, 38(3), 165–180. <https://doi.org/10.1080/1051144X.2019.1575043>
- Cubitt, S., & Cartwright, L. (2018). *Practices of looking: An introduction to visual culture*. Oxford University Press.
- Curtis, D. (1987). *Introduction to visual literacy*. Prentice Hall.
- Danielsson, K., & Selander, S. (2021). Multimodal texts in disciplinary education: A comprehensive framework. *Springer International Publishing*. https://doi.org/10.1007/978-3-030-63960-0_3
- DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. Sage publications.
- Domínguez Romero, E., & Bobkina, J. (2021). Exploring critical and visual literacy needs in digital learning environments: The use of memes in the EFL/ESL university classroom. *Thinking Skills and Creativity*, 40, 100783. <https://doi.org/10.1016/j.tsc.2020.100783>
- Donaghy, K., & Xerri, D. (2017). The image in ELT: An introduction. In K. Donaghy & D. Xerri (Eds.), *The image in English language teaching* (pp. 1–13). ELT Council.
- Dressen-Hammouda, D., & Wigham, C. R. (2022). Evaluating multimodal literacy: Academic and professional interactions around student-produced instructional video tutorials. *System*, 105, 102727. <https://doi.org/10.1016/j.system.2022.102727>
- Elkins, J. (2009). *Visual literacy*. Routledge.
- Felten, P. (2008). Visual literacy. *Change: The Magazine of Higher Learning*, 40(6), 60–64. <https://doi.org/10.3200/CHNG.40.6.60-64>
- Gates, S. (2004). Visual literacy in science and its importance to pupils and teachers. In A. Peacock & A. Cleghorn (Eds.), *Missing the meaning: The development and use of print and non-print text materials in diverse school settings* (pp. 223–237). Palgrave.
- Gnach, A., Weber, W., Engebretsen, M., & Perrin, D. (2022). *Digital communication and media linguistics*. Cambridge University Press.
- Goin, P. (2001). Visual literacy. *Geographical Review*, 91(1–2), 363–369. <https://doi.org/10.1111/j.1931-0846.2001.tb00491.x>

- Group, & T. N. L. (2000). A pedagogy of multiliteracies: Designing social futures. In B. Cope & M. Kalantzis (Eds.), *Multiliteracies: Literacy learning and the design of social futures* (pp. 9–36). Routledge.
- Hall, J. K. (2018). *Essentials of SLA for L2 Teachers: A transdisciplinary framework*. Routledge.
- Hattwig, D., Bussert, K., Medaille, A., & Burgess, J. (2013). Visual literacy standards in higher education: New opportunities for libraries and student learning. *portal: Libraries and the Academy*, 13(1), 61–89. <https://doi.org/10.1353/pla.2013.0008>
- Hermans, P., & Schönau, D. (2022). A plea for visual inquiry. *Journal of Visual Literacy*, 41(3–4), 191–200. <https://doi.org/10.1080/1051144X.2022.2132624>
- Hinkin, T. R. (2005). Scale development principles and practices. In R. A. Swanson & E. F. Holton (Eds.), *Research in organizations: Foundations and methods of inquiry* (pp. 161–179). Berrett-Koehler Publishers Inc.
- Höglund, H. (2022). The heartbeat of poetry: Student videomaking in response to poetry. *Written Communication*, 0(0), 07410883211070862. <https://doi.org/10.1177/07410883211070862>
- Huynh, T.-N., Lin, C.-J., & Hwang, G.-J. (2022). Learner-generated material: The effects of ubiquitous photography on foreign language speaking performance. *Educational Technology Research and Development*, 70(6), 2117–2143. <https://doi.org/10.1007/s11423-022-10149-1>
- Jewitt, C., & Oyama, R. (2001). Visual meaning: A social semiotic approach. In T. Van Leeuwen & C. Jewitt (Eds.), *Handbook of visual analysis* (pp. 134–156). Sage Publications Ltd.
- Kárpáti, A., & Schönau, D. (2022). Introduction to the special issue on the Common European Framework of Visual Competency. *Journal of Visual Literacy*, 41(3–4), 171–177. <https://doi.org/10.1080/1051144X.2022.2132619>
- Katsampoxaki-Hodgetts, K. (2024). Graphical abstracts' pedagogical implications: Skills & challenges in visual remediation. *English for Specific Purposes*, 73, 141–155. <https://doi.org/10.1016/j.esp.2023.10.006>
- Kress, G. (1999). "English" at the crossroads: Rethinking curricula of communication in the context of the turn to the visual. In G. Hawisher & C. Selfe (Eds.), *Passions, pedagogies, and 21st century technologies* (pp. 66–88).
- Kress, G. (2000). Multimodality: Challenges to thinking about language. *TESOL Quarterly*, 34(2), 337–340. <https://doi.org/10.2307/3587959>
- Kress, G. (2003). *Literacy in the new media age*. Routledge.
- Kress, G. (2009). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.
- Kress, G., & Van Leeuwen, T. (2021). *Reading images: The grammar of visual design*. Routledge.
- Lee, C. (2024). Using wordless picturebooks to promote bilingual students' translanguaging practices. *Journal of Research in Childhood Education*, 38(1), 123–144. <https://doi.org/10.1080/02568543.2023.2193258>
- Lim, F. V., Toh, W., & Nguyen, T. T. H. (2022). Multimodality in the English language classroom: A systematic review of literature. *Linguistics and Education*, 69, 101048. <https://doi.org/10.1016/j.linged.2022.101048>
- Lim, J., & Kessler, M. (2024). Multimodal composing and second language acquisition. *Language Teaching*, 57(2), 183–202. <https://doi.org/10.1017/S0261444823000125>
- Matusiak, K. K. (2020). Studying visual literacy: Research methods and the use of visual evidence. *IFLA Journal*, 46(2), 172–181. <https://doi.org/10.1177/0340035219886611>
- Mayer, R. E. (2002). Multimedia learning. In *Psychology of learning and motivation* (Vol. 41, pp. 85–139). Academic Press. [https://doi.org/10.1016/S0079-7421\(02\)80005-6](https://doi.org/10.1016/S0079-7421(02)80005-6)
- Merchant, G. (2012). Mobile practices in everyday life: Popular digital technologies and schooling revisited. *British Journal of Educational Technology*, 43(5), 770–782. <https://doi.org/10.1111/j.1467-8535.2012.01352.x>
- Messaris, P., & Moriarty, S. (2005). Visual literacy theory. In K. L. Smith, S. Moriarty, K. Kenney, & G. Barbatsis (Eds.), *Handbook of visual communication: Theory, methods, and media* (pp. 481–502). Routledge.
- Miller, J. (2014). The fourth screen: Mediatization and the smartphone. *Mobile Media & Communication*, 2(2), 209–226. <https://doi.org/10.1177/2050157914521412>
- Moline, S. (2023). *I see what you mean: Visual literacy K-8*. Routledge.
- Newman, M., & Ogle, D. (2019). *Visual literacy: Reading, thinking, and communicating with visuals*. Rowman & Littlefield Publishers.
- Ng, I. C. L. (2006). Photoessays in the teaching of marketing. *Journal of Marketing Education*, 28(3), 237–253. <https://doi.org/10.1177/0273475306291468>
- Romero, E. D., & Bobkina, J. (2017). Teaching visual literacy through memes in the language classroom. In K. Donaghy & D. Xerri (Eds.), *The image in English language teaching* (pp. 59–70). ELT Council.
- Schönau, D., Kárpáti, A., Kirchner, C., & Letsiou, M. (2020). A new structural model of visual competencies in visual literacy: The Revised Common European Framework of Reference for Visual Competency. *Literacy, Preliteracy and Education*, 4(3), 57–71.
- Shrestha, N. (2021). Factor analysis as a tool for survey analysis. *American Journal of Applied Mathematics and Statistics*, 9(1), 4–11.
- Sireci, S. G. (1998). Gathering and analyzing content validity data. *Educational Assessment*, 5(4), 299–321. https://doi.org/10.1207/s15326977ea0504_2
- Stefanenko, T., & Kupavskaya, A. (2012). Developing cross-cultural competence. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 941–944). Springer.
- Stokes, S. (2002). Visual literacy in teaching and learning: A literature perspective. *Electronic Journal for the Integration of Technology in Education*, 1(1), 10–19.
- Thomson, T. J., & Uddin, S. (2023). Contemporary ways of seeing: Exploring how smartphone cameras shape visual culture and literacy. *Journal of Visual Literacy*, 42(4), 269–286. <https://doi.org/10.1080/1051144X.2023.2281163>
- Tudini, V., & Liddicoat, A. J. (2024). Technology-mediated discourse and second language research. In B. P. M. T. Prior (Ed.), *The Routledge handbook of second language acquisition and discourse* (pp. 297–310). Routledge.
- UNESCO. (2023, February 2). *What you need to know about literacy*. <https://www.unesco.org/en/literacy/need-know>
- Van Leeuwen, T. (2005). *Introducing social semiotics*. Psychology Press.
- Wasilewska, M. (2017). The power of image nation: How to teach a visual generation. In K. Donaghy & D. Xerri (Eds.), *The image in language teaching* (pp. 43–50). ELT Council.

- Winstanley, L., Thompson, J. J., & Tan, S. H. S. (2024). Transformative pedagogy and visual literacy: Reframing art and design student perspectives on sustainability with illustrated infographics. *Journal of Visual Literacy*, 43(2), 73–94. <https://doi.org/10.1080/1051144X.2024.2350240>
- Xueting Ye, S., & Shi, J. (2023). Investigating the potential of changing the smartphone system language to L2 for facilitating vocabulary learning and motivation. *Language Teaching Research*, 0(0), 13621688221145565. <https://doi.org/10.1177/13621688221145565>
- Yenawine, P. (1997). Thoughts on visual literacy. In J. Flood, S. Heath, & D. Lapp (Eds.), *Handbook of research on teaching literacy through the communicative visual arts* (pp. 845–846). Macmillan Library Reference.
- Yeşilel, D. B. A. (2022). Utilizing mobile technology to improve writing skill. In G. Yangın-Ekşi, S. Akayoglu, & L. Anyango (Eds.), *New directions in technology for writing instruction* (pp. 147–167). Springer International Publishing. https://doi.org/10.1007/978-3-031-13540-8_8

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.