



FunQ: Measuring the fun experience of a learning activity with adolescents

Gabriella Tisza¹ · Panos Markopoulos¹

Accepted: 11 February 2021 / Published online: 13 March 2021
© The Author(s) 2021

Abstract

Researchers and practitioners in learning sciences, educational technology and child-computer interaction often argue that fun is an essential element of learning. Therefore, researchers in the above fields aim to explore how learning activities could be made more enjoyable in order to facilitate engagement in the learning process and to improve the learning outcomes. Despite such wide interest, there has been little systematic effort to define and measure fun. The herein introduced research aims to (a) define the term “fun” and (b) to create a tool for the reliable measurement of it. In the first study testing the initial item pool 75 students ($\mu_{\text{age}} = 11.78$); in the think-aloud study testing the comprehensibility of the items six 11-year-old children and in the final validation study, 128 students ($\mu_{\text{age}} = 12.15$) participated. We applied a deductive scale development approach. For the model testing, CFA was used and second-order latent variable models were fitted. In this paper, we conceptualize the term of fun and introduce the final 18-item version of the FunQ that consists of six dimensions (*Autonomy, Challenge, Delight, Immersion, Loss of Social Barriers and Stress*) and bears with the appropriate validity and reliability measures ($\omega_{\text{overall}} = 0.875$ and $\omega_{\text{partial}} = 0.864$; RMSEA = 0.052 and SRMR = 0.072). We contribute with (a) a review of the literature regarding the concept of fun, (b) a conception of fun as a multidimensional theoretically motivated concept, (c) a multidimensional instrument for assessing experienced fun—the FunQ—and (d) a psychometric evaluation of the proposed instrument. FunQ is put forward as a reliable and much-needed addition to the current palette of instruments.

Keywords Experienced fun · Learning · Adolescents · Scale · Validation

Fun is important in several aspects of life. Next to being a defining element of leisure and play, it is increasingly understood to play a key role in learning, work and social interactions. Researchers and developers in the fields of educational technology and child-computer interaction often inject fun elements into their systems just like educationalists who often aspire to make learning activities enjoyable. Research has demonstrated that the promise of fun has an inviting effect (Long 2007), that fun increases engagement with learning technologies and with learning activities (Iten and Petko 2016; Long 2007; Rambli et al. 2013; Vieira and da Silva 2017) and having fun has a positive effect on learning outcomes (Chan et al. 2019; Elton-Chalcraft and Mills 2015;

Long 2007; Lucardie 2014; Rambli et al. 2013). Additionally, neuroscience provides evidence at the biochemical level for the positive effects of fun in the learning environment and on learning (Willis 2007).

Despite that “fun” is often mentioned, the concept behind the term and its measurement are not always clearly defined. The Cambridge dictionary¹ defines “fun” as being an informal expression for *pleasure, enjoyment or entertainment*, and as *behaviour or activities that are not serious; games or jokes*. In the academic literature, the terms “fun” and “enjoyment” are frequently used interchangeably (Elton-Chalcraft and Mills 2015; Fowler 2013; Iten and Petko 2016; Long 2007; Mellecker et al. 2013; Romero 2014). In other cases, these notions stand next to each other as close relatives, however, complementary to some degree (Dismore and Bailey 2011). This also invites a scientific debate concerning the distinction between the two concepts (Dismore and Bailey 2011). Nonetheless, in recent years, it is usual that papers discussing

✉ Gabriella Tisza
g.tisza@tue.nl

¹ Department of Industrial Design, Eindhoven University of Technology, 5612 AE Eindhoven, The Netherlands

¹ <https://dictionary.cambridge.org/dictionary/english/fun>, Retrieved: 18 March, 2019

fun simply do not elaborate on what their understanding of the concept of fun is (Itten and Petko 2016; Knowles et al. 2017; Tokuhisa et al. 2015). In our conceptualization, we understand “enjoyment” as a term that describes positive emotions, while we consider “fun” as a more extensive, nuanced and complex notion, which is yet more difficult to grasp and define. Specifically regarding children, we agree with the viewpoint of Bengoechea et al. (2004) that children are generally seeking positive experiences, and they label those as “fun”.

Despite the growing interest towards measuring the fun experience—with special regard to the relation to learning—currently, there is a lack of reliable measurement tools. Where they exist (mostly in the field of human-computer interaction), they rather measure product liking (acceptance or preference) with young, preliterate children (Fun Toolkit, Read 2008; Fun Semantic Differential Scales, Yusoff et al. 2011; This or That, Zaman et al. 2013). Or in case of adults, the most widely known instruments (EGameFlow, Fu et al. 2009; UES, O’Brien et al. 2018; GEQ, Poels et al. 2013) measure game enjoyment and engagement or the gaming experience along several dimensions. Remarkably, there is a gap in research with adolescents (age 11–18). The aim of the current study is twofold. First, it aims to investigate the notion and the meaning behind the term “fun”. Second, it aims to create a tool for the measurement of the experienced fun that is psychometrically and theoretically sound, comprehensive yet parsimonious, practical and child appropriate, specially developed for adolescents and can be used in the learning environment across various fields of research.

The paper is structured in three main parts. In the first part, we argue why experiencing fun during learning is important. We review the related literature in order to define and summarize the web of related terms. Thereafter, we discuss the methodological challenges of developing questionnaires for children, and we introduce existing tools for measuring fun, including their suitability for our target audience and scope of research (adolescents in the learning environment), and their possible pitfalls.

In the second part, we describe the development of the FunQ. We start with the construction of the initial item pool, we follow with the think-aloud study results; then, we finish with introducing the final version of the instrument, including its psychometric properties. The final version of the FunQ consists of 18 items along six dimensions (*Autonomy, Challenge, Delight, Immersion, Loss of Social Barriers and Stress*) and bears with the appropriate validity and reliability measures ($\omega_{\text{overall}} = 0.875$ and $\omega_{\text{partial}} = 0.864$; RMSEA = 0.052 and SRMR = 0.072).

In the third and last part, we discuss our research findings, the possible applications of the developed measurement tool, and summarize our scientific contribution.

Theoretical Background

Fun and Learning

Research into understanding the role of fun in the learning process is found mainly in the areas of serious games and gamification in education (Chu et al. 2017), thus concerns mainly the non-formal and informal learning environment. In these fields, there is growing evidence in support of the importance of fun for learning; we review this body of research below.

Bisson and Luckner (1996) elaborated on the pedagogical benefits of fun. They saw fun as a powerful tool to enhance motivation and create a safe learning environment. They summarized that fun is beneficial as (a) it evokes intrinsic motivation, (b) it facilitates the suspension of the social reality, (c) it reduces stress and (d) it creates a state of *relaxed alertness* where “learners feel safe to take risks, be creative, make mistakes, and most importantly, keep trying” (Bisson and Luckner 1996, p. 111).

Rambli et al. (2013) stated that fun and interactive learning are one of the most powerful pedagogical factors which could yield to create an interactive and engaged learning environment (Rambli et al. 2013) and they added that this environment facilitates the memorization procedure of learners while keeping their attention, which ultimately enhances learning. Based on the PISA test ($N > 400,000$ 15-year-old students), Ainley and Ainley (2011) concluded that the sense of fun and excitement has a huge importance for science learning.

While investigating the learning outcomes of an educational game, Long (2007) found that 87.5% of the participants joined the activity in the first place because of the promise of fun (*Fun in programming games*) and it improved the skills of 79.8% of the participants. Moreover, she found that computer games “lead to positive results in long-term learner retention by improving learning interest and more focused attention because the students enjoy the approach” (Long 2007, p. 280)—although she used the terms “fun” and “enjoyment” interchangeably. She also demonstrated that fun has a significant effect on the learning effort.

In a recent systemic review on emotions in design-based learning (DBL), Zhang et al. (2020) found that design-based learning had an overall positive effect on students’ interest and motivation to learn. Accordingly, enjoyment was among the most frequently mentioned emotions in DBL; however, this review also reflected that fun and enjoyment are barely distinguished or measured separately.

Vieira and da Silva (2017, p. 130) stated that “fun is an important element of life because it satisfies curiosity and fosters learning”. They encouraged designers to “make their artefacts fun” in order to stimulate users to use them. In their understanding, fun consisted in attention, flow, immersion and emotion. Tews and Noe (2019) even went further and said

that “fun is an important component of high quality learning experiences” (Tews and Noe 2019, p. 226).

Chan et al. (2019) investigated the role of perceived fun in a collaborative learning scenario and the learning performance while using personal response systems (PRSs). Their results suggested that “the level of fun students experienced using PRSs was found to promote collaborative learning and learning performance” (Chan et al. 2019, p. 99).

Iten and Petko (2016) studied whether fun playing an educational game was a predictor for learning success. They used the terms “fun” and “enjoyment” interchangeably. They found that the experienced enjoyment and flow during the game had a significant effect on gaining motivation, increasing interest in the subject matter and upon choosing to play the game again. However, their study could not demonstrate any association between the experienced enjoyment and the learning gains, which is in contrast with previous findings. And, they questioned “whether ‘fun’ and ‘enjoyment’ are adequate constructs to grasp meaningful motivational processes in serious game experiences” (Iten and Petko 2016, p. 161). They referred to other authors who proposed instead “student engagement” to analyze positive emotions when learning with serious games. Similarly, Sim et al. (2006) did not find significant correlation neither between the observed nor the reported fun and the learning outcome.

Controversially, Tews et al. (2017) found that fun had a significant impact on informal learning in the working environment (i.e. learning from others and learning from non-interpersonal sources). They stressed that “researchers should not necessarily focus on fun as a unidimensional construct” (Tews et al. 2017, p. 52). Additionally, their findings suggested that the managers’ support for fun had a significant influence on learning (learning from oneself) as well.

Similarly with adults, but in the learning environment, Lucardie (2014) in her qualitative research found that “both adult learners and their teachers also believed that fun and enjoyment impacted on adults learning and they were able to articulate the role that fun plays in adult learning programs” (Lucardie 2014, p. 445).

Elton-Chalcraft and Mills (2015) summarized their study results as follows: “Learning which is enjoyable (fun) and self-motivating is more effective than sterile (boring) solely teacher-directed learning” (Elton-Chalcraft and Mills 2015, p. 482). This finding is supported by Aoki et al. (2004) who investigated how the education of children with type-1 diabetes could be improved. They developed three edutainment tools and tested them. Their findings suggested that children patients found the games fun (compared to the researchers’ previous study on traditional learning methods), 91.4% of the respondents showed more interest toward the edutainment method, and more than 60% of them found that this approach would be useful as an initial education for type-1 diabetes children. They thus concluded that “edutainment systems could have a significant potential for healthcare education especially for children” (Aoki et al. 2004, p. 859).

Willis (2007) wrote about the neuroscience of joyful education. “When students are engaged and motivated and feel minimal stress, information flows freely through the effective filter in the amygdala and they achieve higher levels of cognition, make connections, and experience ‘aha’ moments. Such learning comes not from quiet classrooms and directed lectures, but from classrooms with an atmosphere of exuberant discovery” (Willis 2007, p. 1). She added that “when classroom activities are pleasurable, the brain releases dopamine, a neurotransmitter that stimulates the memory centers and promotes the release of acetylcholinem, which increases focused attention” (Willis 2007, p. 2). Additionally, she claimed that despite “some schools have unspoken mandates against these valuable components of the classroom experience” (Willis 2007, p. 3), no neuroimaging or brain wave analysis data exist that would demonstrate any downshifting effect of “joy”—a term she used interchangeably with “fun”—in the classroom.

Thus, in sum, previous research provides a growing support for the view that fun has positive effects on learning. However, the above-introduced studies do not define and measure fun, and moreover, none of these studies is a controlled experiment which would compare the effects of introducing fun elements versus not. Only one study—of Iten and Petko (2016)—comes close by studying the effects of enjoyment but in that study, no distinction is made between fun and enjoyment. In order to be able to make such claims precise regarding the effect of fun on learning, or to evaluate such activities, we need to define clearly what fun is and have a reliable instrument to measure it.

Characteristics of Fun

Bisson and Luckner (1996) synthesized earlier scholarly attempts to define fun into four characteristics that were inherent to it. They argued that “fun” is a *relative, situational, voluntary experience* and *natural/essential* to all human beings. With relative and situational, they meant that fun depends on many factors, e.g. what one finds fun is not necessarily fun for another, nor is it certainly fun on another day. Fun is voluntary as “to experience fun one must consciously or unconsciously accept to feel good, to relax, to let go and to let the situation be perceived as enjoyable” (Bisson and Luckner 1996, p. 109). Fun cannot be “forced”, so, for an activity to be perceived as fun, the participation must be intrinsically motivated.

Glasser (1986) argued that fun is one of the five most essential human needs and emphasized its importance while learning and especially for child development. He stated that fun “is like a catalyst that makes anything we do better and worth doing again and again” (Glasser 1986, p. 28). Accordingly, Read et al. (2002) discussed *returnance* as a facet of fun in child-computer interaction; in this context, returnance meant the desire to do an enjoyable activity again and again.

Based on a 3-year-long study on attitudes towards physical education, Dismore and Bailey (2011) argued that the

meaning attributed to the concept of fun changed in merit as children approached teenage years. While for younger children (7–11 years), fun was a critical factor for an activity to be enjoyable; teenagers (11–14 years) described fun in terms of a learning challenge rather than in relation to hedonic responses while playing games (Dismore and Bailey 2011). Along with the aforementioned properties of fun, its stress-reducing effect has to be mentioned as well (Caine and Caine 1991).

Fun and Play

During childhood, “fun” and “play” often concur. Despite that the two concepts are closely related, they should be distinguished from each other. Sutton-Smith (2011, p. 3) defined play as “an activity that is voluntary, intrinsically motivated, fun, incorporates free will/choices, offers escape, and is fundamentally exciting”. Gajadhar, de Kort and IJsselstein considered play as an “intrinsically motivated, physical or mental leisure activity that is undertaken only for enjoyment or amusement and has no other objective” (Gajadhar et al. 2008, p. 105). Especially in early childhood, “fun” and “play” are overlapping notions. However, as indicated already (Dismore and Bailey 2011), with adolescence, the hedonistic character of fun that is purely present during play gives way to challenge, which is a well-established dimension of fun among adults (Fu et al. 2009; IJsselstein et al. 2008; Ryan et al. 2006).

Intrinsic Motivation and Social Aspect

Bisson and Luckner argued that the promise of fun “can motivate learners to engage in activities with which they have little or no previous experience” (Bisson and Luckner 1996, p. 110). Therefore, fun is not only an experience but it can be itself a strong, intrinsic motivating factor to encourage children to try new challenges. Already back in the 1980s, Malone and Lepper (Malone 1981; Malone and Lepper 1987) emphasized the importance of intrinsic motivation that according to their theory, it could be evoked by the optimal level of challenge, curiosity and fantasy. Bisson and Luckner (1996) also suggested that the combination of “fun” and “play” can act as a catalyst to eliminate inhibiting factors inherent to our socialization. In their opinion, “the more genuine and intense the fun is, the greater the suspension of reality will be. Consequently, fun can transform social insecurity into trust and camaraderie, and a restrictive self-image into the freedom of expression” (Bisson and Luckner 1996, p. 110). This property of fun is closely related to the concept of *Flow*.

Flow

The concept, or rather the experience of Flow, was defined by Csikszentmihalyi (1990) as an optimal experience, where the following characteristics were present: (a) an

intrinsically rewarding experience, (b) a loss of reflective self-consciousness, (c) a distorted experience of time, (d) an intense and focused concentration on the present, (e) a merge of actions and awareness, (f) an optimal balance between challenge and skills, (g) a sense of control over the situation and (h) clear goals and immediate feedback. Considering the definitions above, it appears that experiencing the fun and being in the psychological state of Flow overlap substantially.

Abbasi et al. (2019) argued based on structural modelling that the theoretical constructs “experience” (a.k.a. Flow) and “engagement” should not be used interchangeably to investigate the subjective experience of video gameplay. Rather, they proposed a model of playful-consumption experience, which consisted of different types of experience (emotional and sensory) and different types of engagement (cognitive, affective and behavioural), and they discussed “enjoyment” as one of the emotional experience factors, which, in their terminology, were used interchangeably with the term “fun”. Thus, they considered enjoyment or fun as a part of the emotional experience.

As Tasci and Ko (2016, p.167) described, “A distorted sense of time, in general, is taken as an indicator of engagement, desire, enjoyment, excitement and thus, having fun; when it feels as if time went more quickly than it actually did, this is a sign of fun and vice versa”. Rodríguez-Ardura and Meseguer-Artola (2017) also stated that when one feels that the activity is going smoothly and is fun, then time flies, and one undergoes a distortion of the temporal experience of time.

Caine and Caine (1991) showed how learning is maximized when combining fun and challenge, which they called a state of *relaxed alertness* and suggested that a major goal for educators should be to challenge students in a natural way so conceptual mapping (i.e. intellectual connections) could happen without evoking a downshifting response. Mellecker, Lyons and Baranowski have also found while evaluating video game design with children that “an engrossing story in which a player faces increasing challenges and can increase skills quickly enough to overcome the challenges, but not so quickly as to get bored by the challenges, appears to provide an important game design structure for enhancing fun or enjoyment” (Mellecker et al. 2013, p. 144). Chu et al. (2017) described how during curriculum-based making activity children had the most positive feelings and got engaged the most when the level of challenge matched their skills. It has been mentioned already how adolescents attributed experiencing fun to challenge (Dismore and Bailey 2011) which corresponds to the *challenge* aspect of the Flow experience. Rodríguez-Ardura and Meseguer-Artola (2017, p. 902) explained the effect of challenge on Flow in terms of the cognitive evaluation theory (Ryan et al. 1983). They suggested that as long as individuals have a psychological need to feel competent, activities that trigger positive challenge can lead to experiencing optimum experienced and intrinsic motivation—because they satisfy the individual’s need for competence.

Distinction Between Young Children and Adolescents

Approaching the childhood from the perspective of cognitive and psychological development, both Piaget and Erikson account for the shift that occurs on the edge of adolescence. According to Piaget (1964), the *formal operational stage* begins approximately at the age of 11–12 and lasts into adulthood when children develop the ability to think about abstract objects and to logically test hypotheses. In the theory of Erikson (1950), stage 5 is approximately between age 12 and 18, during which children search for a sense of self and personal identity exploring their own personal values, beliefs and goals. Without further delving into the characteristics of the teenagers, it can be summarized that during this age children's identity is formed and the way they understand the world changes. They begin to shape their opinion and they learn how to express it as well. Their cognitive abilities—such as memory capacity, language skills and concentration span—are approaching quickly the level of an adult's. De Leeuw (2011, p. 15) argued that the “age of 11 is seen as a turning point in memory capacity when children appear to function as well as adults”. From the beginning of this age, therefore, they are less prone to the typical response biases that are common for younger children (see section [Attention Span](#)). Ultimately, this means that when they are asked about their opinion, the answers will be more differentiated than in younger ages (this is also shown by Read 2008, 2012; Read and MacFarlane 2006) and are generally more valid (Mellor and Moore 2014). Moreover, Dismore and Bailey (2011) showed that the meaning and the content of the concept of “fun” altered during childhood. This shift was found to be around age 11 as well, which is in synchrony with the psychological and cognitive changes while approaching adolescence. On the basis of these arguments, the definition and measurement of fun will address children over the age of 11.

Surveying Children

In designing measurement tools especially for children, their competencies and differences to adults have to be taken into account. Hall et al. (2016) emphasized that generally, children preferred Likert-type scales over similar simple response items and that free-recall questions were useful especially in spoken surveys (Read and MacFarlane 2006). De Leeuw (2011) found that in general, the older the child, the more reliable the answers will be and that children were better informants on topics directly related to them such as their feelings and other subjective phenomena. Furthermore, de Leeuw (2011, p. 6) stated that “Below the age of 7 children do not have sufficient cognitive skills to be effectively and systematically questioned” and added that individual (semi-) structured interviews were more suitable than questionnaires for children between 7 and 12 (see also Bell 2007). Mellor and

Moore (2014) found that children below the age of 12 had difficulties in answering questions about abstract concepts such as their own behaviours, bodily states or emotional states. They related it to the theory of Piaget about the formal operational stage of development. Regarding adolescents, de Leeuw (2011) suggested that questionnaires could be used for adults; however, there should be special attention devoted against ambiguity in item wording. Therefore, using simple language and formulating items as exactly as possible is a must and ensuring language appropriateness by readability testing is highly recommended.

Read (2008) and de Leeuw (2011) discussed the challenges of designing measurement tools for children, and Mellor and Moore (2014) wrote specifically about the use of Likert-type scales with children. The main caveats which can jeopardize the reliability of surveys involving children are discussed below.

Attention Span

Children's attention span (or sustained attention) is crucial for directly measuring children (i.e. not observing their behaviour). Attention span is defined as the time a person is able to selectively attend to relevant information, such as listening to a teacher and persisting on a task (McClelland et al. 2013, p. 315). Scientists have been studying children's attention span for a long time. In a literature review dating back to the 1950s, Moyer and von Haller Gilmer (1954) reviewed nineteen studies which found that the attention span of young children ranged from 1 to 25 min. Additionally, they found that the attention span was lower in a group situation and that there was a difference “between work tasks, such as reading, and the activities in which a child engages in playing with a toy” (Moyer and von Haller Gilmer 1954, p. 466). Moreover, Sousa (2011) suggested that motivation had an effect on the attention span, and Bradbury (2016) even called this effect crucial.

Within the field of educational psychology, a generally accepted and referred rule of thumb for the length of the average student's attention span is 10 to 15 min during lectures (Goss Lucas and Bernstein 2005; McKeachie and Svinicki 2006, and Wolvin 1983 in Davis 2009). Although formulas are available for approximating the length of the attention span per age, no supporting empirical evidence exists. Lin et al. (1999) showed that the sustained attention developed between ages 6 and 15. Nonetheless, since they have measured the sustained attention by the Continuous Performance Test (CPT), they only report on the hit and false alarm rates for the evaluation of sustained attention and did not provide the length of the attention span in minutes.

Controversially to the generally accepted 10–15 min, a study conducted at Microsoft (2015) claimed that the average attention span was only 12 s in 2000 and 8 s in 2013, and it is ever decreasing. Although the validity of these numbers has been contested (Bradbury 2016), it appears that people

nowadays, and especially the younger generation, get distracted easily and hold their attention for less time than was the case in the past.

To safeguard the reliability of answers, it is important to consider children's attention span adjusting the length of the inspection to the child's attention span. Based on the above, a survey—that is not a particularly engaging task for a child—should not require sustained attention by adolescents for more than 10–15 min.

Bias

When working with children, the risk of introducing bias is high and different types of bias can be manifested compared to those concerning adults. The most common bias types concerning children are discussed below.

Suggestibility pertains to the influence of the researcher on the way the respondent encodes, stores, retrieves and reports events. This effect is due to a range of social and psychological factors. *Social desirability bias* is when the respondent provides the answer that s/he thinks the examiner asking the question wants to hear. *Satisficing* is a tendency of the respondent to select a good enough option, instead of the very best one. In the case of surveys, this phenomenon could be manifested as giving a superficial response that appears to be reasonable but without thoroughly considering all answer possibilities. *Acquiescence bias* is the tendency of the respondent to agree or respond positively. *Extreme responding* is the type of response bias when the respondent mainly selects the most extreme options/answers available. *Straight-lining* is the tendency of the respondent to provide answers in a way that the responses form a line or rather a visual pattern. Extreme responding is a sort of straight-lining; however, meanwhile, in extreme responding, it can be assumed that the respondent reads the question and considers the responses; in case of straight-lining, this assumption cannot be made.

Around the age of 11, the suggestibility of children decreases while the importance of peers increases. Therefore, *peer pressure* can be a serious issue with early adolescents (12–16 years) (de Leeuw 2011). However, contrary to adults, the *item non-response* appears not to be a problem with children and adolescents (Bell 2007). That is, the error size in responses by children and adolescents is approximately stable across different conditions and not dependent on the content of the question. This is assumed to be in relation to their cognitive abilities, namely that they cannot fully apply an optimizing strategy (Borgers 2003); therefore, they will not skip difficult questions. It has, however, a downside. The difficult or vague questions will not be indicated by a missing value pattern, but the quality of those responses remains doubtful. Therefore, the importance of simple and short questionnaire items is stressed and the application of think-aloud interviews to check whether any item is problematic is highly recommended.

Earlier research has shown that children are particularly prone to the above-described bias types (Bell 2007; de Leeuw 2011; Hall et al. 2016; Read 2008; Read and MacFarlane 2006) to very different degrees for different ages, which has to be considered when developing measurement instruments for any specific age group.

Existing Measurement Tools

Methods for evaluating fun derive primarily from the domain of human-computer interaction and especially child-computer interaction, where fun is seen as an essential component of children's interactive experience whether they approach technology as users (e.g. of an application or consumer device), learners or players (Markopoulos et al. 2008).

There are a few existing measurement tools that have been designed to gather opinions on the “funness” of an experience or product. Where these exist, they target either young children or adults. Some studies report the use of a survey or a list of questions to be asked from children within the narrow scope of the study; however, they are not intended for further use nor are they validated. A list including the most-known tools for measuring (aspects of) fun is shown in Table 1.

The This or That (Zaman et al. 2013) method examines preference. Despite being a validated measurement tool, it is constrained by its comparative structure: it is only suitable when measuring the preference of one product/experience over the other, which is particularly suitable for its targeted age group of 2 to 7.

The Fun Semantic Differential Scales (Yusoff et al. 2011) is a measurement tool for evaluating games with nursery-aged children based on choosing between photos of a child expressing different emotions (*love-do not know-hate*). While it has been shown to work well for the target age group, it has not been psychometrically validated, and it addresses fun as a unidimensional construct and is not sufficiently refined for teenagers.

The Fun Toolkit (Read 2008) is a set of tools that targets a wide age range up until teenage to measure the “funness” (Smileyometer) and preference of products (Fun sorter and Again-again table). However, it handles fun as a unidimensional construct, and it faces a problem that younger children tend to use mainly the higher values of the Smileyometer. Despite being widely used, it has not been psychometrically validated.

The Five Degrees of Happiness (Hall et al. 2016) was introduced to address the extreme response bias of the Smileyometer discussed above. Its target audience is children between age 9 and 11, and like its predecessor handles fun as a unidimensional construct and has not been validated psychometrically. Furthermore, the emphasis on positive emotions makes it less suitable for assessing less pleasant experiences that might include frustration or disappointment.

A recent study in this domain proposed a list of Likert scales for the evaluation of a game and attitudes towards

Table 1 Measurement tools for evaluating preference/engagement/experienced fun.

Tool's name	Year	Age	Validated	Internal consistency [†]
This or That (Zaman et al. 2013)	2013	2–7 yrs	Yes	Yes
Fun Semantic Differential Scales (Yusoff et al. 2011)	2011	3–5 yrs	No	No
Fun Toolkit (Read 2008)	2008	4 yrs–teenagers inclusive	No	No
Five degrees of happiness (Hall et al. 2016)	2016	9–11 yrs	No	No
List of statements for the Evaluation of- and attitudes towards learning games (Iten and Petko 2016)	2016	10–13 yrs	No	Yes
Physical Activity Enjoyment Scale (Kendzierski and DeCarlo 1991)	1991	Undergrad. students	Yes	Yes
Player Experience of Need Satisfaction (PENS) (Ryan et al. 2006)	2006	Undergrad. students	Yes (Johnson et al. 2018)	Yes
Game Experience Questionnaire (GEQ) (Poels et al. 2007)	2007	Undergrad. students	Yes	Yes
Playful-consumption experience questionnaire (Abbasi et al. 2019)	2019	16–19 yrs	No	Yes
EGameFlow (Fu et al. 2009)	2009	Undergrad. students	Yes	Yes
FUN scale for understanding the hedonic value of a product (Tasci and Ko 2016)	2016	Adults	Yes	Yes
User Experience Scale (UES) and UES-SF (O'Brien et al. 2018)	2018	Adults	Yes	Yes
EmoForm (Zhang et al. 2019)	2019	13–14 yrs	No	Yes

[†] Whether any internal consistency measures (e.g. Cronbach's alpha or Omega) are published

learning games (Iten and Petko 2016) which however has not been validated. Despite being suitable for teenagers and measuring multiple dimensions, those dimensions are not linked to fun and refer to the serious game rather than to the personal experience (i.e. *How is the game* instead of *How do one feels while playing the game*).

This limitation is avoided in the Physical Activity Enjoyment Scale (Kendzierski and DeCarlo 1991) which measures the personal experience of a physical activity, rather than the activity as such. However, it does not help conceptualize enjoyment which is measured as a unidimensional construct across bipolar scales.

The PENS (Ryan et al. 2006), UES and UES-SF scales (O'Brien et al. 2018) are validated measurement tools, made for adults and have a strong focus on the evaluation of games (usability, aesthetics, novelty, intuitive controls, in-game competence, etc.) rather than on the personal experience (flow, intrinsic motivation, etc.). This limits their applicability in different contexts and does not contribute to our purpose of defining and measuring fun as a psychological construct.

The GEQ (Poels et al. 2007) has the focus on how one feels while playing a game and measures enjoyment as a multidimensional construct; however, it has been validated only with adults in a gaming environment, which limits its applicability in different contexts. Besides, given that the scale is designed for adults, its vocabulary is quite advanced, so it is questionable whether children would be able to comprehend and rate the scale reliably.

In the same domain, a recent study (Abbasi et al. 2019) proposed a Playful-consumption experience questionnaire for the assessment of consumer video game engagement for

adolescents. However, they measured enjoyment as a unidimensional construct and as a subdimension of emotional experience, and the questionnaire has not yet been validated.

The EGameFlow (Fu et al. 2009) measures the enjoyment of an e-learning game across the dimensions of Csikszentmihalyi's Flow theory; thus, it equates enjoyment to the Flow experience, and it is validated for adults. Given the nature of the scale, its usability in different contexts with different ages is limited.

The FUN scale (Tasci and Ko 2016) handles fun as a multidimensional construct; however, it is validated to measure the fun value of a touristic destination as a product among adults, which is reflected in its vocabulary. Additionally, the focus of the scale is to evaluate whether a place, a hotel or a restaurant is fun and not to assess the personal experience.

The EmoForm (Zhang et al. 2019) is a recently developed tool for the assessment of emotions during design-based learning. While the instrument is designed for adolescents, it examines various emotions; however, it does not examine fun only enjoyment across a single item, and it has not yet been validated.

From this review of earlier work, we can conclude that there is currently no psychometrically validated inventory targeting adolescent respondents that is theoretically grounded and that treats fun as a multidimensional construct. The necessity of having multiple dimensions was not only shown by the theoretical review but it helps to conceptualize and define fun rather than treating it as an opaque descriptor or an umbrella term. The present study introduces an instrument designed to fill this gap by providing a tool that is psychometrically and theoretically sound, comprehensive yet parsimonious, practical and

adolescent appropriate, and can be used in the learning environment across various fields of research.

For the development of the FunQ, only a fraction of the referred scales was relevant. It is important to mention that several items of different measurement tools overlap with each other (e.g. items measuring the Flow experience, the perceived competence, the enjoyment). Selected questionnaire items from the EGameFlow (Fu et al. 2009), the Evaluation of- and attitudes towards learning games list (Iten and Petko 2016), the FUN scale (Tasci and Ko 2016), the GEQ (Poels et al. 2007), the Physical Activity Enjoyment Scale (Kendzierski and DeCarlo 1991) and the Intrinsic Motivation Inventory (Ryan 1982) have been included in the initial item pool of FunQ, albeit, rephrased to be adolescent appropriate and to reflect a personal experience instead of evaluating the activity (e.g. *I had fun* instead of *It's [the activity] a lots of fun*). Additionally, the pool of items was extended by further ones that reflect the underlying factors, and the adopted items were organized in the FunQ by the factorial structure proposed in this paper and not by the dimensions of the original instrument.

Development of the FunQ

The development of the FunQ consisted of four main phases applying a deductive scale development approach (Adèr et al. 2008). During the first phase, initially, the dimensions of the instrument were constructed based on the above-described theories. Then, a pool of possible items was created according to the previously defined dimensions and based on the cited theories and existing measurement tools. Second, the initial item pool of the instrument was tested with 75 children. Consequently, a comprehensive yet parsimonious model was created. Third, we conducted think-aloud interviews with six children to assess possible pitfalls relating to the questionnaire items. Fourth, the questionnaire was administered with another 150 children and the validity and reliability measures were calculated.

Methodology of the Survey Design

In the development of the FunQ, we paid attention to a number of key issues as follows:

First, to gauge relevant abstract concepts—such as challenge, fun, flow, stress and autonomy—questions were formulated by asking the emotional and behavioural reflections of these concepts. Then, statements were derived as responses to the questions based on the underlying theories. For example, *How do I feel when I am enjoying an activity?—I feel delighted* and *How do I feel when I'm in Flow?—I feel that time flies*. Then, the emotional and behavioural reflections were filtered, and the wording was adapted to be child

appropriate, e.g. *I feel delighted* was transformed into *I feel good* and *I feel happy*. This approach aligns with the recommendation by de Leeuw (de Leeuw 2011) that the items should be worded to focus on how the child felt during the activity. It is not only easier for children to identify themselves in such items but also this phrasing avoids asking them to judge the activity, thus reducing the risk of social desirability bias, which could arise if they had to evaluate an activity designed by the teacher or another adult. Additionally, we can expect they will be more likely to use the full range of the scale rather than the extremes of the scale. This latter issue has been extensively addressed by previous studies (Hall et al. 2016; Read 2012; Read and MacFarlane 2006; Read 2008) on scales developed specifically for children.

Second, the questionnaire uses contra-indicative statements to measure factors that could indicate one is not having fun. The sentences within these factors are phrased in a non-negating, thus positive way; however, their content is anticipated to be contra-indicative for experiencing fun (e.g. During the activity I felt bad). Such factors are the experienced stress and tension, which are presented in the questionnaire by the initial factors *Fear of Damage*, *Pressure* and *Stress*. Having contra-indicative items among the statements of the questionnaire is intended to prevent—or highlight—*acquiescence bias* and *straight-lining* (see section [Bias](#)) and thus serve as a control for the reliability of the answers.

Third, the questionnaire items were phrased very briefly and in simple language, so that children would find them easy to comprehend and evaluate. Language appropriateness was checked by several measures, which from the Flesch Reading Ease (Flesch 1948) score is 77.9 and the Flesch-Kincaid grade level (Kincaid et al. 1975) is 3.7 for the initial—50 item—version of FunQ, and 84.7 and 2.7 respectively for the 18-item final version. This indicates that the text of the 50-item questionnaire is respectively *fairly easy to read* and is understandable for an average end-of-third-grade student (age 9), and the text of the 18-item final version is *easy to read* and is understandable for an average end-of-second-grade student (age 8). This is in agreement with de Leeuw's suggestion (de Leeuw 2011) that the readability level of items should be about two grades lower than the target group.

Fourth, the appearance of the survey was created by considering the specialties of the target group. Therefore, based on previous findings (Bernard et al. 2001), the text was presented in Comic Sans type with 12 pt. size, which was found to be the most preferred font type and size among 9–11-year-old children. Besides, the items were highlighted with alternating colours—so that it makes easier to keep track of the responses—and additionally, a colourful design was created to make the questionnaire inviting for children (see Fig. [Appendix B](#)). The idea of adding cartoons to the design was considered but was abandoned for fear of the questionnaire appearing too childish for adolescents.

Fifth, there was special attention devoted to the type of response format. Based on the findings of Mellor and Moore (2014) and de Leeuw (2011), it was decided to use a 5-point Likert scale where the points are based on words that reflect the frequency of behaviour/thoughts (i.e. never/rarely/sometimes/often/all the time).

Sixth, the questionnaire was designed so that the response time should stay within the anticipated average concentration span of 10–15 min. The questionnaire items once finalized were randomly mixed.

Construction of the FunQ

The factorial structure of the FunQ was created by adopting a deductive scale development approach (Adèr et al. 2008) following similar steps as in previous research (Di Malta et al. 2020; Polo et al. 2018). That is, the factors were established strictly on the above-referred theories. Based on those theories, we made the assumption that the experienced fun is a multidimensional construct. The importance and the interrelation of the concerning concepts such as control over the activity (Csikszentmihalyi 1990), challenge (Chu et al. 2017; Csikszentmihalyi 1990; Dismore and Bailey 2011; Fu et al. 2009; IJsselsteijn et al. 2008; Mellecker et al. 2013; Ryan et al. 2006), enjoyment (Bisson and Luckner 1996; Glasser 1986; Read et al. 2002), engagement and immersion (Caine and Caine 1991; Csikszentmihalyi 1990; Tasci and Ko 2016), intrinsic motivation (Bisson and Luckner 1996; Csikszentmihalyi 1990; Glasser 1986; Malone 1981; Malone and Lepper 1987), social connectivity (Bisson and Luckner 1996; Csikszentmihalyi 1990) and stress (Caine and Caine 1991) have been argued in the earlier sections. Based on this reasoning, the following factors were established that covered all the referred theories and were expected to be defining for the experienced fun: *Autonomy*, *Challenge*, *Delight*, *Fear of Damage*, *Immersion*, *Loss of Social Barriers*, *Pressure* and *Stress*.

Then, the referred frequently used measurement tools were scrutinized whether they measure any of the dimensions of the FunQ. Consequently, some items from other measurement tools (Fu et al. 2009; Iten and Petko 2016; Kendzierski and DeCarlo 1991; Poels et al. 2007; Ryan 1982; Tasci and Ko 2016) were considered to be taken into the FunQ. For this, the above-detailed protocol for rephrasing and adjustment was followed. Thereafter, the number of questionnaire items was further expanded by items based on the emotional and behavioural reflections of the factor defining concepts while keeping the number of items limited with regard to the attention span of the target respondent population. The item pool was evaluated and adjusted in several consecutive steps according to topic experts' recommendations.

In the initial version of the FunQ, the *Experienced Fun* is measured across eight dimensions. The *Autonomy* factor (4 items) measures whether the child experienced control over the

activity. The *Challenge* factor (10 items) assesses whether the child felt challenged during the activity. The *Delight* factor (9 items) targets the positive emotions experienced during the activity. The *Fear of Damage* factor (4 items) aims to control whether the child experienced fear of hurting someone or causing damage. The *Immersion* factor (8 items) intends to indicate whether the child immersed in the activity by losing the sense of time and space. The *Loss of Social Barriers* factor (4 items) aims to monitor the social connectivity of the child. The *Pressure* factor (5 items) investigates whether the child experiences his/her own participation as voluntary or as obligatory. And, the *Stress* factor (5 items) measures the negative emotions experienced during the activity. In total, the initial version of the questionnaire consisted of 50 items which from 16 were reverse statements (see Appendix A for the items and Appendix B for the design of the questionnaire). The items were evaluated by the children on a 5-step Likert-type scale.

Analysing the Structure of Initial FunQ Item Pool

In order to assess the fit of the FunQ for its aimed purpose, the initial 50-item version of the FunQ was administered to children after they visited an interactive exhibition about the Dutch Delta Works. Based on the statistical analyses, a comprehensive yet parsimonious model was created which contains 18 items across 6 factors. Some items of the final model were slightly adjusted according to the following think-aloud study (see section [Think-Aloud Evaluation of Initial Item Pool](#)), which slightly modified 18-item version of the FunQ was used for the third study.

Think-Aloud Evaluation of Initial Item Pool

Besides statistically testing the fitness of the FunQ, we conducted think-aloud interviews (Ericsson and Simon 1993) with six children, for which we used the initial 50-item version of the instrument. The think-aloud interview is a commonly used method for assessing participant's thought processes especially when confronted with a new situation or artefact. During the interview, the interviewee is asked to verbalize his/her thoughts on the subject of the usability testing while being actively engaged with it. With the think-aloud interviews, we aimed to get an insight (a) whether are problematic or misunderstandable items and (b) whether the children have the same understanding of the questionnaire items as it was intended by the researchers. In our case, the procedure was as follows. To start with, the interviewer explained the method and gave examples to the child of what is expected from him/her. Thereafter, the child was asked to read the FunQ items aloud and verbalize any thoughts that came into his/her mind. The role of the interviewer in this situation was mainly to observe and, if needed, to facilitate the verbalization and eventually to ask for clarification, applying a more relaxed

approach to think-aloud than the recommendations of Erikson and Simon, in order to create a friendly and safe environment for the child participants (Markopoulos et al. 2008).

Psychometric Properties of FunQ

As a final step in the herein described study, we collected further data from 150 children visiting a museum with their school to test the validity and reliability of the final 18-item version of FunQ on a new data set.

Statistical Analyses and Measures

The think-aloud interviews were analyzed qualitatively (see section [Think-Aloud Evaluation of Initial Item Pool](#)). For the assessment of the psychometric properties of the instrument, the statistical analyses and measures are detailed below.

We applied confirmatory factor analysis (CFA) and second-order hierarchical latent variable modelling. Our choice for CFA is supported by the deductive scale development methodology we followed. That is, the FunQ the factors were established strictly on the above-referred theories, and according to previous papers (Hu and Li 2015; Suhr 2006), CFA is the appropriate choice when “the researcher uses knowledge of the theory, empirical research, or both, postulates the relationship pattern a priori and then tests the hypothesis statistically” (Suhr 2006, p. 1.).

Internal Consistency

For measuring internal consistency, Cronbach’s alpha (α) and the Omega (ω) coefficients were calculated. These statistics indicate whether the items measure the same underlying construct. Despite Cronbach’s alpha is the most widely known internal consistency measure, it has been the subject of considerable criticism (Dunn et al. 2014; Morrison et al. 2017; Peters 2014; Revelle and Zinbarg 2009). Thus, we adopt the Omega coefficient as the main indicator for internal consistency as it has been proven to be more reliable (Peters 2014; Revelle and Zinbarg 2009). Given that there is a wide variety for the acceptable internal consistency values (starting from 0.45 (Taber 2018)) for the reported study, we regarded the Omega values above 0.6 acceptable (Hair Jr et al. 2014; Tasci and Ko 2016).

Model Fit

For assessing model fit (whether the factorial structure found in the data is in agreement with the proposed one), we considered a variety of model fit indexes and addressed parameter estimates and their magnitudes. Hu and Bentler (1999) suggested to rely on a combination of indexes that have different measurement properties (e.g. CFI and SRMR). We used the

selected indexes listed below based on the recommendations of Jackson, Gillapsy and Purc-Stephenson and Kline (Jackson et al. 2009; Kline 2015).

- χ^2 value. The χ^2 value is a general, commonly used maximum likelihood approximation for the overall model fit. It tests whether the model implied covariance matrix differs significantly from the measured values. However, it is known that the χ^2 value is affected by the sample size and is mostly significant when $N > 75$.
- *Comparative Fit Index*. The Comparative Fit Index (CFI) is an incremental fit index which ranges from 0 and 1, with higher values indicating better model fit. The cutoff value for the CFI proposed by Hu and Bentler (1999) is 0.95 or higher, which indicates a good model fit.
- *Root Mean Squared Error of Approximation*. The Root Mean Squared Error of Approximation (RMSEA) is a parsimony-based index. The index value typically ranges between 0 and 1, but higher values than 1 are also possible. An index value of 0 indicates a perfect model fit. When the p -value is ≥ 0.05 ; then, the hypothesis of close fit is justified. The cutoff value for the RMSEA proposed by Hu and Bentler (1999) is 0.06 or lower, which indicates a good model fit.
- *Standardized Root Mean Squared Residual*. The Standardized Root Mean Squared Residual (SRMR) is an absolute fit index. The index value ranges between 0 and 1, where 0 marks the perfect fit. Thus, the lower the value, the better the fit is. The cutoff value for the SRMR proposed by Hu and Bentler (1999) is 0.08 or lower, which indicates a good model fit.

For modelling, second-order hierarchical latent variable models were fitted. For the data analysis, we used the RStudio 1.1.453 (RStudio Team 2016) software, for modelling the lavaan package (Rosseel 2012), and the semTools package (Jorgensen et al. 2018) to calculate Cronbach’s alpha and Omega coefficients.

Results

Analysing the Structure of Initial FunQ Item Pool

Data

Data were collected at the beginning of October 2018 from 75 students from the first year of a Dutch secondary school with English speaking specialization (39 boys, 33 girls, 3 missing, $\mu_{age} = 11.78$, $SD_{age} = 0.45$). Consent was attained according to the Dutch regulations: both parents and the child had to sign the informed consent form. The questionnaire was administered on paper after the children attended to Deltapark Neeltje Jans, an interactive

Table 2 Statistics of the final 18-item model on the first data set. The internal consistency coefficients and the standardized factor loadings of the factors on *Experienced Fun*

Factor	Cronbach's alpha	Omega	Standardized factor loading on <i>Experienced Fun</i>	<i>P</i> -value
<i>Autonomy</i>	0.755	0.770	0.763	0.007
<i>Challenge</i>	0.525	0.477	0.719	< 0.001
<i>Delight</i>	0.801	0.808	0.996	<0.001
<i>Immersion</i>	0.408	0.488	1.022	<0.001
<i>Loss of Social Barriers</i>	0.632	0.647	0.577	0.001
<i>Stress (R)</i>	0.862	0.863	0.804	<0.001
<i>Overall</i>	0.867	0.896	-	-

exhibition and information center about the Delta Works, and designed and built their own dams. The data is coming from three groups.

Missing Data

The proportion of missing values both for the whole sample and the questionnaire items is 1.8%. Since the normality of the data cannot be assumed, we used the non-parametric test of homoscedasticity to check whether data is missing completely at random (MCAR). The test resulted in a nonsignificant *p*-value ($p = 0.259$); thus, it was assumed that the values are missing completely at random. Hence, to handle missing data, full information maximum likelihood (FIML) estimation was used.

The Initial 50-Item Pool

To begin with, the construct validity of the proposed factorial structure was assessed by means of confirmatory factor analysis (CFA) and second-order latent variable modelling. The contra-indicative items and factors that are marked with an (R) (see Appendix A) were reversed for the data analysis and for reporting the results. For details about the calculation of the coefficients, see the referred package (Jorgensen et al. 2018). The computation for overall internal consistency ($\omega_{\text{overall}} = 0.916$) at the first-order level and for the second-order factor *Experienced Fun* ($\omega_{\text{partial}} = 0.909$) revealed a high value of Omega coefficient.

The Final 18-Item Model

Since we aimed to create a comprehensive and parsimonious model, the factor loadings of the first-order factors and the

second-order factor *Experienced Fun* were examined. First, it was investigated whether all proposed factors contribute equally well to the second-order factor *Experienced Fun*. The analysis revealed that the *Fear of Damage* factor had no significant effect on the *Experienced Fun* (standardized factor loading = 0.027, $p = 0.891$); therefore, it was removed from the model. Then, the number of the questionnaire items was reduced based on the factor loadings. In case the standardized factor loading of an item was < 0.3, it was considered not to be substantial (Hair Jr et al. 2014) for the given factor and in several consecutive steps, the non-substantial items were removed. Additionally, based on the modification indexes and factor covariances provided by the lavaan package, the *Autonomy* and the *Pressure* factors were merged. During the model fitting process, the internal consistency of the modified factors and the model fit was continuously monitored.

Comparing the final 18-item model to the initial 50-item pool, the *Fear of Damage* factor was completely removed as it appeared not to be related to the second-order factor *Experienced Fun*. Additionally, the *Pressure* and *Autonomy* factors were combined into one factor that measures the free choice/voluntary participation of the child, named *Autonomy*. The final 18-item version of FunQ including the factor loadings is presented in Fig. 1 and Appendix A. The descriptive statistics of the items are presented in Appendix C.

Table 2 summarizes the statistics of the final 18-item model. The internal consistency of the majority of the remaining factors is above the cutoff value ($\omega > 0.6$) that indicates that the items are measuring the same underlying constructs. Despite the internal consistency of the *Challenge* ($\omega = 0.477$) and *Immersion* ($\omega = 0.488$) factors that is below the cutoff value, they both appear to have a significant effect on the *Experienced Fun* ($p_{\text{Challenge}} < 0.001$, $p_{\text{Immersion}} < 0.001$) and a

Table 3 Statistics of the final 18-item model on the first data set. The internal consistency coefficients of *Experienced Fun* as a second-order factor

	Omega at level 1 [†]	Omega at level 2 [‡]	Partial Omega [§]
<i>Experienced Fun</i>	0.822	0.924	0.888

[†] The proportion of the second-order factor explaining the total score

[‡] The proportion of the second-order factor explaining the variance at the first-order factor level

[§] The proportion of observed variance explained by the second-order factor after partialling the uniqueness from the first-order factor

Table 4 Fit indexes of the final 18-item model on the first data set

	χ^2 value	CFI	RMSEA	SRMR
Reference value (Hu and Bentler 1999)	-	> 0.95	< 0.06	< 0.08
Final 18-item model	207.507; df=129, $p<0.01$	0.857	0.090	0.086

standardized factor loading well above the 0.3 margin (0.719 and 1.022 respectively); therefore, the factors were kept in the model. In fact, the *Immersion* factor has a standardized factor loading above 1, which is unusual, however, acceptable, suggesting high correlations among the factors (Jöreskog 1999), that is desirable for second-order latent variable modelling.

The standardized factor loadings of the first-order variables on the second-order variable *Experienced Fun* in the final model are all above the cutoff value of 0.3; therefore, they are considered substantial.

The internal consistency of the second-order factor *Experienced Fun* is presented in Table 3. The Omega values are found to be above the cutoff value ($\omega > 0.6$). This finding suggests that the *Autonomy*, *Challenge*, *Delight*, *Immersion*, *Loss of Social Barriers* and *Stress* factors measure with high reliability the same underlying construct, the *Experienced Fun*.

The model fit indexes of the final 18-item model are introduced in Table 4. Despite the borderline values, given the relatively small sample size, the sufficient factor loadings, p -values and internal consistency values, in order to prevent hard fitting to the data, we decided to adhere to this model and test it on a new, larger data set.

Think-Aloud Evaluation of Initial Item Pool

Data

The think-aloud interviews were conducted on 14 February 2019 at an international school in the Netherlands with six 11-year-old children after participating in a playful learning activity during which they prototyped a robot. Consent was attained according to the Dutch regulations: both parents and the child had to sign the informed consent form. Strengthening further the voluntary character of the participation in the interview, from the fifteen children who delivered the consent form complete, six who were willing were invited for the think-aloud interviews.

Analysis

For the think-aloud interviews, the initial 50-item pool of the FunQ was used for the sake of completeness. For the evaluation of the interviews, and thus, for the usability testing of the FunQ, the following aspects were considered:

Table 5 FunQ items and the usability issue based on the think-aloud interviews

Item	Usability issue
[D4] <i>I want to do the activity again.</i>	“Well, I think the answer would be never because why would I do the same thing, make the same robot again? But maybe I could change it. So the answer would be rarely, but if the activity was the same but I got to make some different things like a different robot, then I would take it as often”
[S6] <i>This was an activity that I could not do very well.</i>	‘could not’ was continuously misread as ‘could’
[P3] <i>I did this activity because I had no choice.</i>	“Yes, I had to, but if I had a choice I’d still do it. But I’ll just have to put all the time”
[A4] <i>I could make some choices about the activity.</i>	Some asked for clarification, some interpreted is as follows: “There could be some things which could be improved”
[I7] <i>I forgot about troubles.</i>	“I do not really have any troubles” “I could not forget about them because I did not have any – so I could say never”
[P1] <i>I felt it was not my choice to do the activity.</i>	‘not’ was continuously skipped while reading aloud and interpreting the item
[I8] <i>I forgot about my daily routine.</i>	“I do not have a daily routine” “I do not have a daily routine – so I’ll say all the time”
[I6] <i>I forgot about my homework.</i>	“we do not have any homework” “no, ‘cause I do not have homework” “yes, because we do not really have homework in this school”

Table 6 Statistics of the 18-item model on the second data set. The internal consistency coefficients and the standardized factor loadings of the factors on *Experienced Fun*

Factor	Cronbach's alpha	Omega	Standardized factor loading on <i>Experienced Fun</i>	<i>P</i> -value
<i>Autonomy</i>	0.578	0.653	0.734	0.006
<i>Challenge</i>	0.434	0.425	0.990	0.002
<i>Delight</i>	0.751	0.766	0.993	< 0.001
<i>Immersion</i>	0.640	0.608	0.885	< 0.001
<i>Loss of Social Barriers</i>	0.611	0.633	0.516	0.006
<i>Stress (R)</i>	0.790	0.791	0.649	< 0.001
Overall	0.843	0.875	-	-

- Misreading of words
- Difficulty with reading the item (when no reading difficulty was observed in general)
- Asking clarification about the item
- Adding comments that imply that the item is not relevant
- Interpreting the item in a way that does not align with the intended meaning

In general, a good usability of FunQ was found. Specifically, the suitability of the design, the language and the scale labelling was justified by the interviews. It appeared that children—even being at the lowest range of the target user age group—went through most of the questions smoothly. Implying that FunQ is user-friendly, the appearance supports the evaluation of the items, which are readable (font type and size) and understandable

Fig. 1 The second-order hierarchical model results of the final 18-item model on the second data set. Standardized factor loadings are shown. All of the related *p*-values are below the 0.05 margin

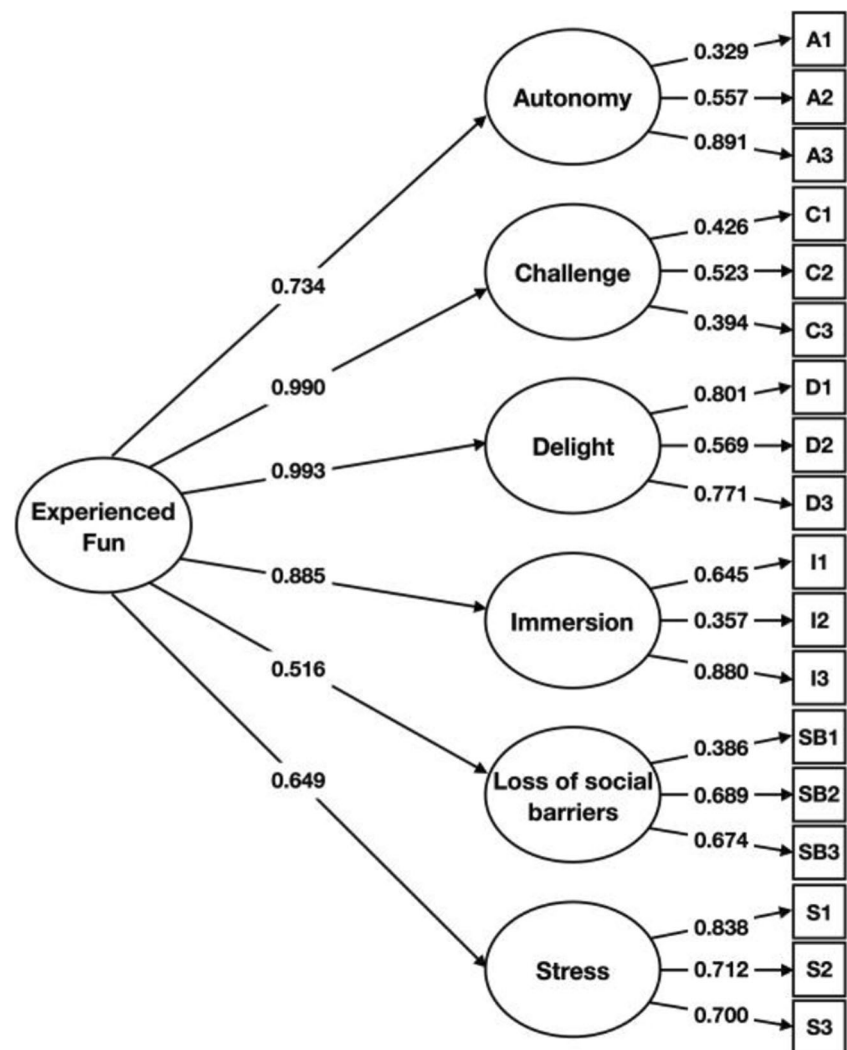


Table 7 Statistics of the 18-item model on the second data set. The internal consistency coefficients of *Experienced Fun* as a second-order factor

	Omega at level 1 [†]	Omega at level 2 [‡]	Partial Omega [§]
<i>Experienced Fun</i>	0.794	0.928	0.864

[†] The proportion of the second-order factor explaining the total score

[‡] The proportion of the second-order factor explaining the variance at the first-order factor level

[§] The proportion of observed variance explained by the second-order factor after partialling the uniqueness from the first-order factor

(language). Also, the used labelling of the steps of the scale (chosen base on the suggestions of de Leeuw (de Leeuw 2011)) appeared to be adequate as it helped children to think back and identify themselves with the statements.

According to the above-established criteria, based on the interviews, eight items emerged as problematic. Those are detailed in Table 5.

The items indicated by the interviews and the statistical analyses of the initial item pool were in approximate overlap, except for the items of the *Fear of damage* factor. None of those items appeared to be problematic during the interviews. Therefore, we can conclude that the items of the *Fear of Damage* factor are understandable and comprehensible to children. Hence, the nonsignificant effect on *Experienced Fun* is not due to the quality of the items.

According to the results of the think-aloud interviews, slight modifications were applied in general to the questionnaire and specific to item D4. Namely, item D4 was slightly modified from *I want to do the activity again* to *I want to do something like this again*. Additionally, since most of the children misread the activity as *this* activity, it was corrected accordingly in the whole questionnaire. Consequently, the questionnaire was used in this modified way during the next data collection and thus in the validation of the FunQ.

Psychometric Properties of FunQ

Data

Data were collected between 8 and 17 March 2019 during the British Science Week at the Science Museum. The reason for collecting the second data set in the UK was to collect responses from native English-speaking adolescents in order to ensure the quality of the instrument. The questionnaire responses were obtained from eight school classes who attained during this period the interactive Wonderlab: The Equinor Gellery program. In total, 150 responses were collected; however, the quality of 22 responses was questionable. They showed signs of typical

response bias (straight-lining; see section [Bias](#)) or many responses were missing (e.g. the second half of the questionnaire) questioning the reliability of those responses. For the sake of data quality, the data of those 22 respondents were completely removed before the analysis started.

The validation was conducted on the data of the remaining 128 respondents (64 boys, 45 girls, 19 missing, $\mu_{\text{age}} = 12.15$, $SD_{\text{age}} = 1.079$). The questionnaire was administered on paper after the children had participated in the activity. Consent was attained across the class teachers according to the British regulations.

Missing Data

The proportion of missing values for the whole sample is 2% and for the questionnaire items is 1.7%. Since the normality of the data cannot be assumed, we used the non-parametric test of homoscedasticity to check whether data is missing completely at random (MCAR). The test resulted in a nonsignificant p -value ($p = 0.093$), thus it was assumed that the values are missing completely at random. Hence, to handle missing data, full information maximum likelihood (FIML) estimation was used.

The Model Fit of the 18-Item Model

To assess the validity of the previously established final 18-item model, we fitted it to the second data set. Table 6 introduces the statistics of the 18-item model on the second dataset; Fig. 1 depicts the model with the related factor loadings. The internal consistency of the majority of the remaining factors is above the cutoff value ($\omega > 0.6$) suggesting that the items are measuring the same underlying constructs. Despite that the internal consistency of the factor *Challenge* ($\omega = 0.425$) is below the cutoff value, it appears to have a strong significant effect (std. factor loading = 0.990, $p = 0.002$) on the *Experienced Fun*; therefore, keeping the factor in the model is justified. The standardized factor loadings of the first-order variables on the second-order variable *Experienced Fun* of the

Table 8 Fit indexes of the 18-item model on the second data set

	χ^2 value	CFI	RMSEA	SRMR
Reference value (Hu and Bentler 1999)	-	> 0.95	< 0.06	< 0.08
18-item model	173.632; df=129, $p=0.005$	0.933	0.052	0.072

final model on the validation data set are all above the cutoff value of 0.3; therefore, they are considered substantial.

The internal consistency of the second-order factor *Experienced Fun* is presented in Table 7. The Omega values are found to be above the cutoff value ($\omega > 0.6$). This finding suggests that the *Autonomy*, *Challenge*, *Delight*, *Immersion*, *Loss of Social Barriers* and *Stress* factors measure with high reliability the same underlying construct, the *Experienced Fun*.

The model fit indices of the 18-item model on the second data set are introduced in Table 8. Evaluating the model fit indices and taking into account the limitations of the χ^2 test, we can conclude that based on the RMSEA and SRMR values, the model fit is sufficient.

Discussion

In recent years, it has become common practice to address “fun” as a *common sense* notion instead of precisely defining the meaning of the concept (Iten and Petko 2016; Knowles et al. 2017; Tokuhiya et al. 2015) making its measurement difficult. Therefore, the aim of the paper was twofold: (a) defining the concept of fun and (b) creating a tool for the multidimensional measurement of the experienced fun that is psychometrically and theoretically sound, comprehensive yet parsimonious, practical and child appropriate, specially developed for adolescents and that can be used in the learning environment across various fields of research. To this end, we have adopted a deductive scale development approach, which is widely used in the field of industrial and organizational psychology (Tay and Jebb 2017). Accordingly, the conceptualization of the construct of fun was theory driven based on a thorough review of literature related to fun. We examined a network of related concepts contributing a theoretically founded conception of fun for our targeted demographic. We concluded that for adolescents to experience an activity as fun they need (a) to feel in control of the activity and be intrinsically motivated for participation (*Autonomy*); (b) to experience an optimal level of the challenge matching their level of skills (*Challenge*); (c) to feel *well* during the activity (*Delight*) and (d) to not feel *bad* (*Stress*, contra-indicative); (e) to be immersed in the activity losing one’s perception of time and space (*Immersion*) and (f) to let go of social inhibitions (*Loss of Social Barriers*). The FunQ is put forward as a tool for testing how a learning activity maps on its different dimensions.

Our conception of fun was tested by the statistical analysis of the created instrument. The final model consists of 18 items across six dimensions. Besides statistically testing the instrument, the comprehensibility and appropriateness for the target age group were checked by the think-aloud interviews, and the questionnaire was adapted accordingly. The final version of the FunQ has been shown to have reliable internal

consistency both at the first- and second-order levels. Since the two data sets (for testing the initial item pool and for validating the final 18-item version) were collected at two different countries (The Netherlands and the UK), from eleven groups of adolescents who participated in three different kinds of learning activities, it is assumed that the revealed model is not activity specific. Additionally, given that the FunQ items are phrased in a general way, we anticipate that the instrument will be applicable in a broad range of different contexts to assess the experienced fun of an activity among adolescents.

The data analysis confirms our initial expectations that fun is a multidimensional construct. Among the dimensions of fun examined it seems that the *Fear of Damage* has no significant effect on whether adolescents experience an activity as fun; however, the existence of the rest of the proposed factors is confirmed with the note that the *Autonomy* and *Pressure* factors were merged as they appeared to measure the two extremes of the same dimension.

With the largest standardized factor loading, the *Delight* factor has the greatest contribution to the *Experienced Fun*. This factor focuses on the positive emotions and the related desires. It sounds natural that fun is a positive experience, and as such, it implies the desire for repetition (Bisson and Luckner 1996; Glasser 1986; Read et al. 2002). This aspect is captured by the *Delight* factor, which our findings indicate as an organic part of the *Experienced Fun*.

To maintain the engagement and therefore to stay in the activity while experiencing it continuously as fun, however, the optimal level of challenge is required. While previous researches with children investigated challenge and fun as separate constructs (Caine and Caine 1991; Chu et al. 2017), our model considers that challenge is a facet of the experienced fun. This idea appears in measurement tools designed for adults (Fu et al. 2009; Poels et al. 2007), however, for adolescents, the association has only been highlighted in relation to physical education activities by the qualitative study of Dismore and Bailey (2011). Our findings suggest that *Challenge* is the second most important factor of the *Experienced Fun*, though it is left for future investigations to establish the suitability of challenge as a dimension of fun for children of different ages.

The *Immersion* factor measures the loss of time and space. When one is deeply engaged, the immersion in the activity happens that leads to the loss of sense of time and space (Csikszentmihalyi 1990; Rodríguez-Ardura and Meseguer-Artola 2017; Tasci and Ko 2016). In the FunQ, this aspect is mapped by the *Immersion* factor which was found to have the third-highest factor loading on the *Experienced Fun*.

The *Autonomy* factor bears the fourth-highest standardized loading and it assesses whether the child feels control over his/her participation as well as the activity itself. As it was summarized above, fun is a voluntary experience (Bisson and Luckner 1996); therefore, intrinsic motivation is seen a key

factor for participation (Mellecker et al. 2013). Additionally, applying the Flow theory (Csikszentmihalyi 1990), it was expected that feeling in control over the situation is related to motivation as well. Our finding supports this theory as the *Autonomy* factor that refers to the experienced control over the participation and the activity itself appeared to have a significant effect on the *Experienced Fun*.

Compared to the research instruments currently used to measure fun, e.g., in the context of evaluations of interactive systems and educational (serious) games, this study also includes contra-indicative items and factors to the construct of *Experienced Fun* to enhance the validity of the tool but also to allow assessing whether the activity that is intended to be fun causes unintentionally any distress to the participants. The antagonism between stress and fun has previously been taken as obvious. Caine and Caine (1991) mentioned the stress-reducing effect of fun, however, without statistically testing it. Our findings provide supportive evidence that negative emotions are contra-indicative for experiencing fun as the effect of the *Stress* factor was found to be significant.

According to Flow theory (Csikszentmihalyi 1990), immersion should result in social barriers to be largely removed. That is, while experiencing the fun and immersing in the activity, the suspension of reality is triggered, which, in turn, leads to loss of self-consciousness. Once the person is less self-conscious, he/she is becoming less engaged with him/herself, is less afraid of rejection, and more open for others, which ultimately results in the breakdown of social barriers. Bisson and Luckner (1996) indicated that in the case of children, the combination of fun and play could act as a catalyst to eliminate inhibitions inherent to our socialization. Our findings support this theory as the *Loss of Social Barriers* factor had a significant contribution to the *Experienced Fun*. In other words, while having fun, children could connect to each other easier than usual.

Regarding the psychometric properties, the internal consistency measures (Cronbach's alpha and Omega) for the second-order factor *Experienced Fun* provide evidence that the questionnaire measures reliably the underlying construct. And, the model fit indices suggest a sufficient model fit. It is therefore proposed that the FunQ is suitable and valid to measure the experienced fun with adolescents. However, the role of challenge on the experienced fun among adolescents is proposed to be further investigated, especially as the internal consistency of the *Challenge* factor did not meet the criterion level ($\omega > 0.6$).

Comparing FunQ to other instruments, FunQ covers a similar ground to the This or That (Zaman et al. 2013), the Fun Semantic Differential Scales (Yusoff et al. 2011), the Fun Toolkit (Read 2008) and the Five Degrees of Happiness (Hall et al. 2016) instruments; however, FunQ is a theoretically founded instrument, which handles fun as a multidimensional construct instead of being unidimensional, and it is designed for- and validated with adolescents instead of young

children. Regarding the PENS (Ryan et al. 2006), the UES and UES-SF (O'Brien et al. 2018), the GEQ (Poels et al. 2007), the Playful-consumption experience questionnaire (Abbasi et al. 2019) and the EGameFlow (Fu et al. 2009) scales and the list of Likert scales for the evaluation of a game and attitudes towards learning games (Iten and Petko 2016), they all designed for the gaming environment; hence, their usability is limited in the learning environment for which FunQ has been created. Additionally, the aforementioned scales mainly target adults and mostly focus on the evaluation of a product or game, in comparison with FunQ, which is designed for adolescents and focuses on the personal experience while being engaged with a learning activity. The FUN scale (Tasci and Ko 2016) is validated to measure the fun value of a touristic destination as a product while the Physical Activity Enjoyment Scale (Kendzierski and DeCarlo 1991), as its name suggests, evaluates a physical activity; thus, they both target a different field than FunQ. Comparing FunQ to the EmoForm (Zhang et al. 2019), they are both designed for adolescents and for the learning environment, but the former focuses on the experienced fun as a multidimensional construct, while the latter investigates a broader range of emotions, handles enjoyment unidimensionally and has not been validated yet. Therefore, we conclude that FunQ is a much-needed instrument measuring fun as a multidimensional construct covering playful learning activities involving adolescents.

In sum, this paper contributes to (a) a review of the literature regarding the concept of fun, (b) a conception of fun as a multidimensional theoretically motivated concept, (c) a multidimensional instrument for assessing experienced fun—the FunQ targets specifically adolescents both in the design, the content and the response format—and (d) a psychometric evaluation and validation of the proposed instrument.

Limitations and Future Work

The herein introduced study is limited to the general population of adolescents in the learning environment. Additionally, as mentioned above, particular attention should be paid to investigate further the role of challenge on the experienced fun for different ages and settings. To further expand the potential of the questionnaire, follow-up studies shall investigate the psychometric properties of the questionnaire for different ages and examine its scope of application: whether the FunQ can be applied to evaluate fun not only in relation to learning but in other activities in which fun can play a useful role, such as participation in experimental studies, child-computer interaction, playful activities and experiences. The validation of the instrument in other languages is already in progress. With this latter, we also aim to investigate the intercultural differences in the concept and experience of fun.

Appendix A

The FunQ questionnaire sorted by the factorial structure and with standardized factor loadings of the concerning factors of the final 18-item model on the second data set. Italic entries indicates the initial items that were removed for the final version

Nr.	Factor	Init. label	Item	Std. factor load.	P- value
1	Autonomy ($\omega=0.653$)	A1	<i>During the activity, I had the sense of controlling the activity.</i>	0.329	-
2		A2	During the activity, I knew what to do.		
3		A3	<i>During the activity, I could do what I wanted.</i>		
4		A4	<i>During the activity, I could make some choices about the activity.</i>		
5		P1 [†]	<i>During the activity, I felt it was not my own choice to do the activity. (R)</i> [‡]		
6	Challenge ($\omega=0.990$)	P2	<i>During the activity, I felt like I had to do the activity. (R)</i>	0.557	0.009
7		P3	<i>I did this activity because I had no choice. (R)</i>		
8		P4	I did this activity because I had to. (R)		
9		P5	I did this activity because I wanted to.		
10		C1	<i>The activity was easy for me.</i>		
11		C2	During the activity, I felt I was good at this activity.		
12		C3	<i>During the activity, I had to concentrate hard.</i>		
13		C4	<i>The activity was difficult for me. (R)</i>		
14		C5	<i>During the activity, I felt challenged.</i>		
15		C7	During the activity, I did something new.		
16	Delight ($\omega=0.993$)	C8	<i>During the activity, I did something I'd never done before.</i>	0.523	0.036
17		C9	During the activity, I was curious.		
18		C10	<i>During the activity, I felt smart.</i>		
19		D1	During the activity, I had fun.		
20		D2	<i>I liked the activity.</i>		
21		D3	<i>I enjoyed doing the activity.</i>		
22		D4	I want to do something like this again.		
23		D5	<i>During the activity, I laughed a lot.</i>		
24		D6	During the activity, I was happy.		
25		D7	<i>During the activity, I had a lot of energy.</i>		
26	Fear of Damage	D8	<i>During the activity I was excited.</i>	0.771	< 0.001
27		C6	During the activity, I was bored. (R)		
28		FD1	<i>During the activity, I was scared of breaking something. (R)</i>		
29		FD2	<i>During the activity, I was afraid of damaging something. (R)</i>		
30		FD3	<i>During the activity, I was afraid of making mistakes. (R)</i>		
31		FD4	<i>During the activity, I was afraid of hurting someone. (R)</i>		
32		I1	<i>During the activity, I was only thinking about the activity.</i>		
33		I2	<i>During the activity, I forgot everything around me.</i>		
34		I3	During the activity, I felt that time flew.		
35		I4	<i>During the activity, I forgot where I was.</i>		
36	Immersion ($\omega=0.885$)	I5	During the activity, I forgot about school.	0.645	-
37		I6	<i>During the activity, I forgot about homework.</i>		
38		I7	<i>During the activity, I forgot about troubles.</i>		
39		I8	<i>During the activity, I forgot about my daily routine.</i>		
40		D9	During the activity, I felt good.		
41		SB1	During the activity, I made new friends.		
42		SB2	During the activity, I talked to others easier than usual.		
43		SB3	During the activity, I felt closer to others more than usual.		
44		SB4	<i>During the activity, I talked with others to whom I had never before.</i>		
45		S1	During the activity, I felt bad. (R)		
46	Stress ($\omega=0.791$)	S2	<i>During the activity, I felt irritated. (R)</i>	0.838	-
47		S3	During the activity, I felt angry. (R)		
48		S4	During the activity, I felt sad. (R)		
49		S5	<i>During the activity, I was anxious. (R)</i>		
50		S6	<i>This was an activity that I could not do very well. (R)</i>		

[†] Label P indicates the initial factor *Pressure*

[‡] (R) indicates a reverse statement


Appendix B

Gender: boy / girl Date: _____ Nr: _____
Age: _____ Activity code: _____
(indicated by administrator)

Hi there,

I would like to ask you to evaluate the activity you've just participated in. There are no good or bad answers. Just let me know how did you experience it. If you have any questions, feel free to ask!

Thank you!

Please indicate with an 

	never	rarely	some-times	often	all the time
I want to do the activity again.	0	0	0	0	0
This was an activity that I couldn't do very well.	0	0	0	0	0
I did this activity because I had to.	0	0	0	0	0
I liked the activity.	0	0	0	0	0
The activity was difficult for me.	0	0	0	0	0
I enjoyed doing the activity.	0	0	0	0	0
I did this activity because I had no choice.	0	0	0	0	0
The activity was easy for me.	0	0	0	0	0
I did this activity because I wanted to.	0	0	0	0	0

(GO TO NEXT PAGE)

Gender: boy / girl Date: _____ Nr: _____
Age: _____ Activity code: _____
(indicated by administrator)

During the activity...


	never	rarely	some-times	often	all the time
I thought I was good at the activity.	0	0	0	0	0
I talked to others easier than usual.	0	0	0	0	0
I was scared of breaking something.	0	0	0	0	0
I felt smart.	0	0	0	0	0
I laughed a lot.	0	0	0	0	0
I felt that time flew.	0	0	0	0	0
I could make some choices about the activity.	0	0	0	0	0
I felt good.	0	0	0	0	0
I felt challenged.	0	0	0	0	0
I had the sense of controlling the activity.	0	0	0	0	0
I forgot about troubles.	0	0	0	0	0
I was afraid of making mistakes.	0	0	0	0	0
I forgot everything around me.	0	0	0	0	0
I felt angry.	0	0	0	0	0
I was bored.	0	0	0	0	0
I felt like I had to do the activity.	0	0	0	0	0
I knew what to do.	0	0	0	0	0
I made new friends.	0	0	0	0	0
I was happy.	0	0	0	0	0
I was nervous.	0	0	0	0	0
I had a lot of energy.	0	0	0	0	0
I felt irritated.	0	0	0	0	0
I forgot where I was.	0	0	0	0	0

(GO TO NEXT PAGE)

Gender: boy / girl Date: _____ Nr: _____
Age: _____ Activity code: _____
(indicated by administrator)

During the activity...

	never	rarely	some-times	often	all the time
I was excited.	0	0	0	0	0
I forgot about school.	0	0	0	0	0
I was afraid of hurting someone.	0	0	0	0	0
I could do what I wanted to.	0	0	0	0	0
I was afraid of damaging something.	0	0	0	0	0
I felt sad.	0	0	0	0	0
I felt it was not my own choice to do the activity.	0	0	0	0	0
I did something I'd never done before.	0	0	0	0	0
I had fun.	0	0	0	0	0
I talked with others to whom I had never before.	0	0	0	0	0
I was only thinking about the activity.	0	0	0	0	0
I felt bad.	0	0	0	0	0
I felt closer to others more than usual.	0	0	0	0	0
I had to concentrate hard.	0	0	0	0	0
I forgot about my daily routine.	0	0	0	0	0
I did something new.	0	0	0	0	0
I was curious.	0	0	0	0	0
I forgot about homework.	0	0	0	0	0

THANK YOU! 

The design of the FunQ (initial 50-item version)

Appendix C

Descriptive statistics and test of normality of FunQ items on the second data set ($N = 128$)

Item	Mean [†]	Standard deviation	Skewness	Kurtosis	Shapiro-Wilk test of normality
During the activity...					
I knew what to do	3.94	1.091	-0.705	-0.371	0.000
I felt I was good at this activity.	3.96	0.991	-0.660	-0.172	0.000
I did something new	3.86	1.293	-0.909	-0.249	0.000
I was curious.	3.54	1.407	-0.602	-0.861	0.000
I had fun.	4.22	1.023	-1.107	0.346	0.000
I was happy.	4.02	0.963	-0.469	-0.976	0.000
I felt that time flew.	3.64	1.329	-0.604	-0.729	0.000
I forgot about school.	3.09	1.635	-0.077	-1.606	0.000
I felt good	4.04	1.053	-1.066	0.769	0.000
I made new friends.	1.72	1.147	1.570	1.468	0.000
I talked to others easier than usual.	3.12	1.462	-0.085	-1.358	0.000
I felt closer to others more than usual.	2.44	1.384	0.389	-1.207	0.000
I felt bad. (R)	1.39	0.858	2.401	5.472	0.000
I felt angry. (R)	1.30	0.772	3.021	9.528	0.000
I felt sad. (R)	1.50	1.005	2.193	4.281	0.000
I did this activity because I had to. (R)	2.42	1.347	0.511	-0.966	0.000
I did this activity because I wanted to.	3.66	1.250	-0.594	-0.566	0.000
I want to do something like this again.	3.70	0.929	-0.212	-0.258	0.000

[†] 1, never; 5, all the time

Acknowledgements We are grateful to the anonymous students and their teachers for their participation in the herein introduced studies. Special thanks to Bethan Ross from the Science Museum Group for her help with the UK data collection.

Funding This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement NO 787476.

Availability of Data and Material

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

Declarations

Ethical Approval At the time of the study, the Eindhoven University of Technology, Faculty of Industrial Design, had no policy for ethical clearance, neither had an ethical committee. Nevertheless, the study was conducted with adherence to the highest ethical standards with regard to existing laws and regulations on personal data, privacy and research data management.

Informed Consent Participation in the study was voluntary and informed consent was obtained from the study participants and their parents/caretakers.

Conflict of Interest The authors declare no conflict of interest.

Disclaimer This paper reflects only the authors' views. The Research Executive Agency (REA) and the European Commission are not responsible for any use that may be made of the information it contains.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abbasi, A. Z., Ting, D. H., Hlavacs, H., Costa, L. V., & Veloso, A. I. (2019). An empirical validation of consumer video game engagement: A playful-consumption experience approach. *Entertainment Computing*, 29(December 2018), 43–55. <https://doi.org/10.1016/j.entcom.2018.12.002>.
- Adèr, H. J., Mellenbergh, G. J., & Hand, D. J. (2008). *Advising on research methods: A consultant's companion*. Huizen, The Netherlands: Johannes van Kessel Publishing.
- Ainley, M., & Ainley, J. (2011). Student engagement with science in early adolescence : The contribution of enjoyment to students' continuing interest in learning about science. *Contemporary Educational Psychology*, 36(1), 4–12. <https://doi.org/10.1016/j.cedpsych.2010.08.001>.
- Aoki, N., Ohta, S., Masuda, H., Naito, T., Sawai, T., Nishida, K., et al. (2004). Edutainment tools for initial education of type-1 diabetes mellitus: Initial diabetes education with fun. *Studies in Health Technology and Informatics*, 107(May 2014), 855–859. <https://doi.org/10.3233/978-1-60750-949-3-855>.
- Bell, A. (2007). Designing and testing questionnaires for children. *Journal of Research in Nursing*, 12(5), 461–469. <https://doi.org/10.1177/17449871079616>.
- Bengoechea, E. G., Streat, W. B., & Williams, D. J. (2004). Understanding and promoting fun in youth sport: Coaches' perspectives. *Physical Education & Sport Pedagogy*, 9(2), 197–214. <https://doi.org/10.1080/1740898042000294994>.
- Bernard, M., Mills, M., Talissa, F., & McKown, J. (2001). Which fonts do children prefer to read online? Usability News, 3(1). Retrieved from <http://usabilitynews.org/which-fonts-do-children-prefer-to-read-online/>
- Bisson, C., & Luckner, J. (1996). Fun in learning: The pedagogical role of fun in adventure education. *The Journal of Experimental Education*, 19(2), 108–112. <https://doi.org/10.1177/105382599601900208>.
- Borgers, N. (2003). Response quality in survey research with children and adolescents: The effect of labeled response options and vague quantifiers. *International Journal of Public Opinion Research*, 15(1), 83–94. <https://doi.org/10.1093/ijpor/15.1.83>.
- Bradbury, N. A. (2016). Attention span during lectures: 8 seconds, 10 minutes, or more? *Advances in Physiology Education*, 40(4), 509–513. <https://doi.org/10.1152/advan.00109.2016>.
- Caine, R. N., & Caine, G. (1991). *Making connections: Teaching and the human brain*. Alexandria, Virginia: Association for Supervision and Curriculum Development. <https://doi.org/10.1038/ni1002-883>.
- Chan, S. C. H., Wan, J. C. L., & Ko, S. (2019). Interactivity, active collaborative learning, and learning performance: The moderating role of perceived fun by using personal response systems. *International Journal of Management Education*, 17(1), 94–102. <https://doi.org/10.1016/j.ijme.2018.12.004>.
- Chu, S. L., Angello, G., Saenz, M., & Quek, F. (2017). Fun in making: Understanding the experience of fun and learning through curriculum-based making in the elementary school classroom. *Entertainment Computing*, 18, 31–40. <https://doi.org/10.1016/j.entcom.2016.08.007>.
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York, New York, USA: Harper & Row.
- Davis, B. G. (2009). *Tools for teaching*. San Francisco, CA: Jossey-Bass. <https://doi.org/10.1186/1472-6920-11-18>.
- de Leeuw, E. D. (2011). Improving data quality when surveying children and adolescents : Cognitive and social development and its role in questionnaire construction and pretesting.
- Di Malta, G., Evans, C., & Cooper, M. (2020). Development and validation of the relational depth frequency scale. *Psychotherapy Research*, 30(2), 213–227. <https://doi.org/10.1080/10503307.2019.1585590>.
- Dismore, H., & Bailey, R. (2011). Fun and enjoyment in physical education: Young people's attitudes. *Research Papers in Education*, 26(4), 499–516. <https://doi.org/10.1080/02671522.2010.484866>.
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399–412. <https://doi.org/10.1111/bjop.12046>.
- Elton-Chalcraft, S., & Mills, K. (2015). Measuring challenge, fun and sterility on a 'phonometre' scale: Evaluating creative teaching and learning with children and their student teachers in the primary school. *Education 3–13*, 43(5), 482–497. <https://doi.org/10.1080/03004279.2013.822904>.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data, rev. ed. protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA, US: The MIT Press.

- Erikson, E. H. (1950). *Childhood and society*. New York, New York, USA: Norton & Company.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233.
- Fowler, A. (2013). Measuring learning and fun in video games for young children: A proposed method. In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13* (p. 639). New York, New York, USA: ACM Press. <https://doi.org/10.1145/2485760.2485879>.
- Fu, F. L., Su, R. C., & Yu, S. C. (2009). EGameFlow: A scale to measure learners' enjoyment of e-learning games. *Computers and Education*, 52(1), 101–112. <https://doi.org/10.1016/j.compedu.2008.07.004>.
- Gajadhar, B. J., de Kort, Y. A. W., & IJsselstein, W. A. (2008). Shared fun is doubled fun: Player enjoyment as a function of social setting. In *Proceedings of Fun and Games Second International Conference Eindhoven* (pp. 106–117). https://doi.org/10.1007/3-540-68339-9_34.
- Glasser, W. (1986). *Control theory in the classroom*. New York, NY: Perennial Library/Harper & Row Publishers.
- Goss Lucas, S., & Bernstein, D. A. (2005). *Teaching psychology: A step by step guide*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Hair Jr, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate Data Analysis* (7th edition). Pearson Education Limited. Retrieved from www.pearsoned.co.uk0A.
- Hall, L., Hume, C., & Tazzyman, S. (2016). Five degrees of happiness: Effective smiley face Likert scales for evaluating with children. *Proceedings of the 15th international conference on interaction design and children*, 311–321. <https://doi.org/10.1145/2930674.2930719>.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>.
- Hu, Z., & Li, J. (2015). The integration of EFA and CFA: One method of evaluating the construct validity. *Global Journal of Human-Social Science: Arts & Humanities -Psychology*, 15(6), 15–19.
- IJsselstein, W., Van Den Hoogen, W., Klimmt, C., de Kort, Y., Lindley, C., Mathiak, K., et al. (2008). Measuring the experience of digital game enjoyment. *Proceedings of Measuring Behavior*, 2008(2008), 88–89. <https://doi.org/10.3758/BRM.41.3.717>.
- Iten, N., & Petko, D. (2016). Learning with serious games: Is fun playing the game a predictor of learning success? *British Journal of Educational Technology*, 47(1), 151–163. <https://doi.org/10.1111/bjet.12226>.
- Jackson, D. L., Gillaspay, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, 14(1), 6–23. <https://doi.org/10.1037/a0014694>.
- Johnson, D., Gardner, M. J., & Perry, R. (2018). Validation of two game experience scales: The Player Experience of Need Satisfaction (PENS) and Game Experience Questionnaire (GEQ). *International Journal of Human Computer Studies*, 118(February), 38–46. <https://doi.org/10.1016/j.ijhcs.2018.05.003>.
- Jöreskog, K. G. (1999). How large can a standardized coefficient be? Unpublished report. Retrieved from <http://www.ssicentral.com/lisrel/techdocs/HowLargeCanaStandardizedCoefficientbe.pdf>
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2018). *semTools: Useful tools for structural equation modeling*. R package version 0.5–1.
- Kendzierski, D., & DeCarlo, K. J. (1991). Physical activity enjoyment scale: Two validation studies. *Journal of Sport and Exercise Psychology*, 13(1), 50–64. <https://doi.org/10.1123/jsep.13.1.50>.
- Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (automated readability index, fog count and Flesch reading ease formula) for navy enlisted personnel. <https://doi.org/10.21236/ADA006655>.
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*(Fourth Edi). The Guilford Press.
- Knowles, C., Harris, A., & Van Norman, R. (2017). Family fun nights: Collaborative parent education accessible for diverse learning abilities. *Early Childhood Education Journal*, 45(3), 393–401. <https://doi.org/10.1007/s10643-016-0801-2>.
- Lin, C. H., Hsiao, C., & Chen, W. (1999). Development of sustained attention assessed using the continuous performance test among children 6–15 years of age. *Journal of Abnormal Child Psychology*, 27(5), 403–412. <https://doi.org/10.1023/A:1021932119311>.
- Long, J. (2007). Just for fun: Using programming games in software programming training and education - a field study of IBM robocode community. *Journal of Information Technology Education*, 6, 279–290. <https://doi.org/10.28945/216>.
- Lucardie, D. (2014). The impact of fun and enjoyment on Adult's learning. *Procedia - Social and Behavioral Sciences*, 142, 439–446. <https://doi.org/10.1016/j.sbspro.2014.07.696>.
- Malone, T. W. (1981). Toward a theory of intrinsically instruction motivating. *Cognitive Science*, 5(4), 333–369. https://doi.org/10.1207/s15516709cog0504_2.
- Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. *Aptitude Learning and Instruction*, 40, 41. [https://doi.org/10.1016/S0037-6337\(09\)70509-1](https://doi.org/10.1016/S0037-6337(09)70509-1).
- Markopoulos, P., Read, J. C., MacFarlane, S., & Hoysniemi, J. (2008). *Evaluating children's interactive products: Principles and practices for interaction designers*. San Francisco, CA: Morgan-Kaufmann. <https://doi.org/10.1016/B978-0-12-374111-0.X0001-1>.
- McClelland, M. M., Acock, A. C., Piccinin, A., Rhea, S. A., & Stallings, M. C. (2013). Relations between preschool attention span-persistence and age 25 educational outcomes. *Early Childhood Research Quarterly*, 28(2), 314–324. <https://doi.org/10.1016/j.ecresq.2012.07.008>.
- McKeachie, W. J., & Svinicki, M. (2006). *McKeachie's teaching tips: Strategies, research, and theory for college and university teachers* (12th ed.). Boston, MA: Houghton-Mifflin.
- Mellecker, R., Lyons, E. J., & Baranowski, T. (2013). Disentangling fun and enjoyment in exergames using an expanded design, play, experience framework: A narrative review. *Games for Health Journal*, 2(3), 142–149. <https://doi.org/10.1089/g4h.2013.0022>.
- Mellor, D., & Moore, K. A. (2014). The use of likert scales with children. *Journal of Pediatric Psychology*, 39(3), 369–379. <https://doi.org/10.1093/jpepsy/jst079>.
- Microsoft Canada. (2015). Attention spans. *Consumer Insights*, 1–52.
- Morrison, T. G., Morrison, M. A., & McCutcheon, J. M. (2017). Best practice recommendations for using structural equation modelling in psychological research. *Psychology*, 08(09), 1326–1341. <https://doi.org/10.4236/psych.2017.89086>.
- Moyer, K. E., & von Haller Gilmer, B. (1954). The concept of attention spans in children. *The Elementary School Journal*, 54(8), 464–466. <https://doi.org/10.1086/458623>.
- O'Brien, H., Cairns, P., & Hall, M. (2018). A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human - Computer Studies*, 112, 28–39. <https://doi.org/10.1016/j.ijhcs.2018.01.004>.
- Peters, G.-J. Y. (2014). The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *European Health Psychologist*, 16(2), 56–69.
- Piaget, J. (1964). Part I: Cognitive development in children: Development and learning. *Journal of Research in Science Teaching*, 2(3), 176–186. <https://doi.org/10.1002/tea.3660020306>.

- Poels, K., de Kort, Y. A. W., & IJsselstein, W. A. (2007). D3.3: Game Experience Questionnaire: Development of a self-report measure to assess the psychological impact of digital games. *Eindhoven*.
- Poels, K., de Kort, Y. A. W., & IJsselstein, W. A. (2013). The Game Experience Questionnaire. Retrieved from <https://pure.tue.nl/ws/portalfiles/portal/21666907>
- Polo, F., Cervai, S., & Kantola, J. (2018). Training culture. *Journal of Workplace Learning*, 30(3), 162–173. <https://doi.org/10.1108/jwl-01-2018-0024>.
- Rambli, D. R. A., Matcha, W., & Sulaiman, S. (2013). Fun learning with AR alphabet book for preschool children. *Procedia Computer Science*, 25, 211–219. <https://doi.org/10.1016/j.procs.2013.11.026>.
- Read, J. C. (2008). Validating the Fun Toolkit: An instrument for measuring children's opinions of technology. *Cognition, Technology & Work*, 10(2), 119–128. <https://doi.org/10.1007/s10111-007-0069-9>.
- Read, J. C. (2012). Evaluating artefacts with children: Age and technology effects in the reporting of expected and experienced fun. *Icmi*, 2012. <https://doi.org/10.1145/2388676.2388727>.
- Read, J. C., & MacFarlane, S. (2006). Using the fun toolkit and other survey methods to gather opinions in child computer interaction. In *Proceeding of the 2006 conference on Interaction design and children - IDC '06* (p. 81). <https://doi.org/10.1145/1139073.1139096>.
- Read, J. C., Macfarlane, S., & Casey, C. (2002). Endurability, engagement and expectations: Measuring children's fun. *Interaction Design and Children*, 2, 1–23. [10.1111.100.9319](https://doi.org/10.1111.100.9319).
- Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, 74(1), 145–154. <https://doi.org/10.1007/s11336-008-9102-z>.
- Rodríguez-Ardura, I., & Meseguer-Artola, A. (2017). Flow in e-learning: What drives it and why it matters. *British Journal of Educational Technology*, 48(4), 899–915. <https://doi.org/10.1111/bjet.12480>.
- Romero, V. (2014). Children's experiences: Enjoyment and fun as additional encouragement for walking to school. *Journal of Transport and Health*, 2(2), 230–237. <https://doi.org/10.1016/j.jth.2015.01.002>.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36.
- RStudio Team. (2016). RStudio: Integrated development environment for R. Boston, MA. Retrieved from <http://www.rstudio.com/>
- Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, 43(3), 450–461. <https://doi.org/10.1037/0022-3514.43.3.450>.
- Ryan, R. M., Mims, V., & Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology*, 45(4), 736–750. <https://doi.org/10.1037/0022-3514.45.4.736>.
- Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion*, 30(4), 347–363. <https://doi.org/10.1007/s11031-006-9051-8>.
- Sim, G., MacFarlane, S., & Read, J. C. (2006). All work and no play: Measuring fun, usability, and learning in software for children. *Computers & Education*, 46(3), 235–248. <https://doi.org/10.1016/j.compedu.2005.11.021>.
- Sousa, D. A. (2011). *How the brain learns* (4th ed.). Corwin, CA: Thousand Oaks.
- Suhr, D. (2006). Exploratory or confirmatory factor analysis ? In *Proceedings of the 31st Annual SAS Users Group International Conference* (p. Paper number 200-31). Cary, NC: SAS Institute Inc.
- Sutton-Smith, B. (2011). The ambiguity of play (1st edition). Harvard University Press. Retrieved from [https://books.google.nl/books?hl=en&lr=&id=AgA8q0TCkElC&oi=fnd&pg=PR5&dq=Sutton-Smith,+B.++\(2001\).+The+Ambiguity+of+Play+\(1st+ed.\).+Harvard+University+Press.&ots=Cpa4B_t1aO&sig=OjfnfsOLacc9fPIIuK5AdniEz04#v=onepage&q&f=false](https://books.google.nl/books?hl=en&lr=&id=AgA8q0TCkElC&oi=fnd&pg=PR5&dq=Sutton-Smith,+B.++(2001).+The+Ambiguity+of+Play+(1st+ed.).+Harvard+University+Press.&ots=Cpa4B_t1aO&sig=OjfnfsOLacc9fPIIuK5AdniEz04#v=onepage&q&f=false)
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>.
- Tasci, A. D. A., & Ko, Y. J. (2016). A FUN-SCALE for understanding the hedonic value of a product: The destination context. *Journal of Travel and Tourism Marketing*, 33(2), 162–183. <https://doi.org/10.1080/10548408.2015.1038421>.
- Tay, L., & Jebb, A. T. (2017). Scale development. In S. Rogelberg (Ed.), *The SAGE Encyclopedia of Industrial and Organizational Psychology* (2nd ed.). CA: Sage: Thousand Oaks.
- Tews, M. J., & Noe, R. A. (2019). Does training have to be fun? A review and conceptual model of the role of fun in workplace training. *Human Resource Management Review*, 29(2), 226–238. <https://doi.org/10.1016/j.hmr.2017.11.003>.
- Tews, M. J., Michel, J. W., & Noe, R. A. (2017). Does fun promote learning? The relationship between fun in the workplace and informal learning. *Journal of Vocational Behavior*, 98, 46–55. <https://doi.org/10.1016/j.jvb.2016.09.006>.
- Tokuhisa, S., Kamiyama, Y., & Tokiwa, T. (2015). Personal, physical, social, and creative contextual design for art education. *Computers in Entertainment*, 11(4), 1–20. <https://doi.org/10.1145/2582193.2633443>.
- Vieira, L. C., & da Silva, F. S. C. (2017). Assessment of fun in interactive systems: A survey. *Cognitive Systems Research*, 41, 130–143. <https://doi.org/10.1016/j.cogsys.2016.09.007>.
- Willis, J. (2007). The neuroscience of joyful education. *Educational Leadership*, 64, 1–4 Retrieved from <http://www.ascd.org/publications/educational-leadership/summer07/vol64/num09/The-Neuroscience-of-Joyful-Education.aspx>.
- Wolvin, A. D. (1983). Improving listening skills. In R. B. Rubin (Ed.), *Improving speaking and listening skills. New directions for college learning assistance*. San Francisco: Jossey-Bass.
- Yusoff, Y. M., Ruthven, I., & Landoni, M. (2011). The fun semantic differential scales. In *Proceedings of the 10th International Conference on Interaction Design and Children - IDC '11* (pp. 221–224). <https://doi.org/10.1145/1999030.1999066>.
- Zaman, B., Vanden Abeele, V., & De Grooff, D. (2013). Measuring product liking in preschool children: An evaluation of the Smileyometer and This or That methods. *International Journal of Child-Computer Interaction*, 1(2), 61–70. <https://doi.org/10.1016/j.ijcci.2012.12.001>.
- Zhang, F., Markopoulos, P., Bekker, T., Schüll, M., & Paule-Ruiz, M. (2019). EmoForm. In *Proceedings of FabLearn 2019* (pp. 18–25). New York, NY, USA: ACM. <https://doi.org/10.1145/3311890.3311893>.
- Zhang, F., Markopoulos, P., & Bekker, T. (2020). Children's emotions in design-based learning: A systematic review. *Journal of Science Education and Technology*, 29(4), 459–481. <https://doi.org/10.1007/s10956-020-09830-y>.

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