Pilot study of an intervention based on an intelligent tutoring system (ITS) for instructing mathematical skills of students with ASD and/or ID

Cécile Mazon, Benjamin Clément, Didier Roy, Pierre-Yves Oudeyer, Hélène Sauzéon

Pilot study of an intervention based on an intelligent tutoring system (ITS) for instructing mathematical skills of students with ASD and/or ID

Cécile Mazon[1,2*], Benjamin Clément[2], Didier Roy[2], Pierre-Yves Oudeyer[2], Hélène Sauzéon[1,2*]

[1]Active Team, U1219, Bordeaux Population Health Inserm-Université de Bordeaux, Bordeaux, France

[2]Flowers Project-team, Inria center of University of Bordeaux, Talence, France

Abstract

Few technology-based interventions have addressed mathematical and numeracy skills of individuals with autism spectrum disorder (ASD). However, children and adolescents with ASD may experience difficulties in mathematical learning, and even be behind their peers at school. Intelligent Tutoring System (ITS) is an innovative way for enhancing teaching programs through learning optimization algorithms coping interindividual differences in the educational curriculum progress. They have been used with success with typically-developed students for levering on the linkage learning-motivation, but remains at a theoretical stage for students with ASD and/or intellectual disability (ID). We conducted a pilot study to assess an ITS-based intervention embedded into the *KidLearn* application for instructing mathematical skills (i.e., adding and subtracting numbers between 0 and 100). Twenty-four secondary school students with ASD and/or intellectual disability, enrolled from special education classrooms, have been recruited: 14 students received the *KidLearn* intervention, while the 10 others received control intervention. Pre-post assessments showed that calculation and motivational measures were significantly improved in *KidLearn* condition compared to the control condition. ITS intervention appeared relevant for calculation learning in the setting of special education.

Keywords: Autism spectrum disorder, intelligent tutoring system, mathematical instruction, special need education.

Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by impairments in communication and social interactions, as well as a particular pattern of restricted, repetitive and stereotyped behaviors, activities and interests (APA, 2013). ASD is characterized by a wide heterogeneity in its clinical presentation, which mainly results from the high variability in the symptom severity and the frequent association with comorbidities such as an intellectual disability (APA, 2013). Individuals with ASD alone have also cognitive particularities, which may interfere with their learning. Hence, ASD may affect students' learning abilities, especially when associated with intellectual or learning disabilities and result in a lesser school performance (Jones, et al., 2009; Oswald, et al., 2016; Wei, et al., 2015).

The academic achievement during elementary school years is one of the decisional factors for the school placement at the secondary level: the greater the academic performance is, the more chances the students have to access to the regular classrooms (Jones, et al., 2009; Kurth & Mastergeorge, 2010; Wei, et al., 2015). The placement in specialized vs. regular classroom depends also on cognitive and adaptive skills of the child (Jones, et al., 2009; Kurth & Mastergeorge, 2010). Consequently, specialized classrooms are often composed with students having an ASD with a comorbid condition, such as intellectual or learning disabilities, which may experience more educational gaps (Fletcher, et al., 2010; Kurth & Mastergeorge, 2010; Wei, et al., 2015). Without intervention targeting this academic learning gap, ASD students in specialized classrooms have less opportunities for inclusion in regular classrooms, and then, for social participation.

Compared to research on clinical outcomes of ASD intervention, the academic performance of students with ASD have received less attention (*e.g.*, Gevarter, et al., 2016; Hart Barnett & Cleary, 2015; Keen, et al., 2016; Knight, et al., 2013; Mazon, et al., 2019; Wong, et al., 2015) and few technology-based interventions have targeted the learning of mathematics and numeracy in students with ASD and/or ID (Knight, et al., 2013; Spooner, et al., 2019). However, about 25% of students with ASD are found to have delays or difficulties in acquiring mathematics and numeracy skills (*e.g.*, Gevarter, et al., 2016; Hart Barnett & Cleary, 2015; Jones, et al., 2009), and even more when there is a comorbid cognitive condition, such as Intellectual Disability (ID) or Learning Disabilities (LD). Such basic skills are however essential for performing a range of daily living activities, and for promoting their independence in adulthood (Fleury, et al., 2014; Keen, et al., 2016). Hence, we explored to what extent an innovative intervention, based on an Intelligent Tutorial System (ITS) that maximize learning progress for each individual during money exchange activities, can increase mathematical learning performance of students with ASD and/or ID.

Mathematical skills in ASD and/or ID

Mathematical abilities vary among individuals with ASD. Previous studies reported large interindividual differences, with profiles from normal performance to major difficulties (Jones, et al., 2009; Oswald, et al., 2016; Wei, et al., 2015). Individuals with the lesser symptom severity often displayed a normal-to-superior level in mathematics, even if they may experience some difficulties related to verbal aspects, abstract reasoning and problem solving (Gevarter, et al., 2016; Oswald, et al., 2016). Some individuals even had a special talent for mathematics, sometimes termed « hypercalculia », but math giftedness is not a hallmark of individuals with ASD (Bullen, et al., 2020; Jones, et al., 2009; Maras, et al., 2019; Oswald, et al., 2016; Wei, et al., 2015). Indeed, the studies exploring mathematical abilities of individuals with ASD reported also a significant proportion of individuals having difficulties in learning mathematics and acquiring numeracy skills, performing at least one standard deviation below their age norm (Bullen, et al., 2020; Chiang & Lin, 2007; Jones, et al., 2009; Morsanyi, et al., 2018; Oswald, et al., 2016; Wei, et al., 2015).

Both IQ and autism symptomatology has been identified as predictors of school achievement (Fleury, et al., 2014; Jones, et al., 2009; Keen, et al., 2016; Wei, et al., 2015), but there are still questions regarding the extent to which either ASD and ID specifically contribute to the delayed mathematical abilities (Jones, et al., 2009; Oswald, et al., 2016; Wei, et al., 2015). The severity of ASD symptoms has been related to the poorer academic performance, including mathematics (Fleury, et al., 2014; Keen, et al., 2016). The two core ASD symptoms, that is social communication impairments and restrictive, repetitive and stereotypic behaviors, appeared to be highly related with school academic achievement (Brennan, et al., 2012; Hammer, et al., 2017; Jahromi, et al., 2013; Morgan, et al., 2015; Zaidman-Zait, et al., 2020). Indeed, social skills are important as the instruction is mainly provided by human teachers, and students with ASD may struggle in capturing information during classroom courses (Fleury, et al., 2014). Students with ASD and ID have often more difficulties in mathematics than students with ASD (Jones, et al., 2009; Oswald, et al., 2016; Wei, et al., 2015). However, ID alone cannot fully account for mathematical learning delays observed in students with ASD, as even ASD student without ID can experience delays in both basic and complex mathematical skills (Bullen, et al., 2020; Chiang & Lin, 2007; Morsanyi, et al., 2018).

A common cognitive feature between ASD and ID is the executive dysfunction, as executive impairments have been documented in individuals with either ASD (Barendse, et al., 2013; Hill, 2004; Kiep & Spek, 2017, Ozonoff & Schetter, 2007), ID (Danielsson, et al., 2010; Panerai, et al., 2014; Roelofs, et al., 2015), or both (Panerai, et al., 2014; Roelofs, et al., 2015). Executive function (EF) is an umbrella term referring to a set of cognitive functions that underlie flexible, self-regulated, and goal-oriented behaviors in novel or complex situations, including inhibition, flexibility, updating, planning, self-regulation, and metacognitive skills (Friedman & Miyake, 2017; Gioia, et al., 2002; Meltzer & Krishnan, 2007). These functions are central for learning, as EF support the information processing in a strategic manner, the problem-solving mechanisms, and contribute to consolidate and generalize acquired skills and knowledge (Meltzer & Krishnan, 2007; for a review of EF in education, see Meltzer, 2010). Impairments in EF are related to learning difficulties, as poorer EF predict lesser learning capabilities and poorer school achievement (Cragg & Gilmore, 2014; Fleury, et al., 2014; Geary, 2011; Oswald, et al., 2016; Samuels, et al., 2016). Mathematical learning appears to be more correlated with the updating/working memory EF component, than with reading achievement (Cragg & Gilmore, 2014; Geary, 2011). Some studies even showed that individuals with ASD and normal-to-high IQ have poorer WM abilities than neurotypical individuals (Barendse, et al., 2013; Hill, 2004; Kiep & Spek, 2017). Also, individuals with ASD have metacognitive difficulties, which reduce their ability to detect and correct their errors (Brosnan, et al., 2016; Maras, et al., 2019).

Both behavioral and digital interventions have less addressed mathematical learning in students with ASD and/or ID than language, communication, and social outcomes (Knight, et al., 2013; Mazon, et al., 2019; Spooner, et al., 2019; Wong, et al., 2015). Among the few existing mathematical learning-related studies, there were mainly researches focusing on the assessment of pedagogical interventions for basic mathematical skills (*e.g.*, adding, subtracting, money and purchasing) in students with ASD and/or ID (Bouck, et al., 2020; Cihak & Grim, 2008; Fletcher, et al., 2010; Hart Barnett & Cleary, 2015; Hsu, et al., 2014; Oswald, et al., 2016; Spooner, et al., 2019). The counting strategies and manipulatives were reported as efficient interventions for mathematical learning (Bouck, et al., 2020; Cihak & Grim, 2008; Fletcher, et al., 2010; Hsu, et al., 2014; Spooner, et al., 2019). When these pedagogical strategies were embedded in digital applications, they have been identified as useful and potentially effective for supporting the learning of students with special educational needs (Fletcher, et al., 2010; Kiru, et al., 2018; Pitchford, et al., 2018; Spooner, et al., 2019).

Digital interventions in education: ITS and their advantages

As in ASD interventional studies, educational research leverages digital progress to support the learning activities of learners. Digital technologies offer promising ways to meet the challenge of responding to the interindividual variability of learners. To this end, intelligent tutoring systems (ITS) have been developed into the artificial intelligence research. Such systems are designed for providing personalized instruction and adaptability to the learners (Kulik & Fletcher, 2016; Ma, 2014; VanLehn, 2011). In other words, ITS provide a way to develop educational programs that are customizable and continuously adapted to the learner’s profile and progress. The efficacy of ITS with neurotypical children has been demonstrated in several studies, and several meta-analyses confirmed that ITS elicited positive effects on learning (Kulik & Fletcher, 2016; Ma, 2014; VanLehn, 2011). In a recent meta-analysis, Kulik & Fletcher (2016) showed that ITS are very effective instructional tools and elicited moderate-to-large effects compared with conventional instruction, either with a traditional computer-based instruction or a human tutor.

Surprisingly, ITS remain little explored in interventions with students with ASD or other neurodevelopmental disorders. They may yet provide an effective way to respond to the high variability in developmental trajectories observed among ASD individuals. Indeed, the main objective of an ITS is to select activities that best meet learners’ needs and difficulties, and to provide adequate feedback (Clément, et al., 2013; Kulik & Fletcher, 2016; Ma, 2014; VanLehn, 2011). Several studies examined the feasibility of teaching students with ASD using ITS on a theoretical basis (Aljameel, et al., 2017; Sarma & Ravindran, 2007): they modeled an ASD student model and performed simulations showing that ITS instruction was feasible with ASD students. The LANA ITS described in Aljameel, et al. (2017) was further explored in an experimental trial with neurotypical children and yielded promising results (Aljameel, et al., 2019). A recent study reported on the architecture of an ITS designed for students with ASD, and planned to test it on an ASD student model (Milani, et al., 2020). Though these studies argue for the relevance of ITS instruction for ASD, they did not carry out experimental trial with students with ASD to assess the impact of ITS instruction in real settings, and very few studies reported such results in the ASD literature. We found another pilot study using ITS for instructing mathematics to students with ASD (Mondragon, et al., 2016): the authors reported preliminary results showing that an ITS that provides a pedagogical agent with accompaniment features, can increase the mathematical performance of students with ASD.

Actually, ITS seek to provide learning environments that adapt to learners' idiosyncrasies and maximize their learning gains. To this end, most recent and promising ITS are developed with learning optimization algorithms. They are designed to select activities in order that maximize students' learning gains, based on their online performance (Ma, 2014; Nkambou, et al., 2010). Clément, *et al.* (2013) have developed an algorithm for activity selection and embedded it in a tablet application named *KidLearn*, which approaches the instruction of arithmetic skills through a series of monetary-exchange activities. Their algorithm, called ZPDES for *Zone of Proximal Development and Empirical Success*, defines an activity space depending on the student's progress. Contrary to many ITS algorithms (Aljameel, et al., 2019; Sarma & Ravindran, 2007), ZPDES is not based on a predefined student model. Instead, ZPDES selects activities based on the actual performance of the student according to a formalization of the concept of "zone of proximal development" (Vygotsky, 1980). The learning activities are then selected to be neither too easy nor too difficult on the basis of information from the whole set of activities and the student actual performance. A study with neurotypical children showed that the ZPDES algorithm positively affects intrinsic motivation to learn and maximizes students' progress in mathematics in comparison with an expert-predefined exercise sequence condition (Clément, 2018; Clément, et al., 2013). ZPDES algorithm has also been applied with success to the therapeutic education of asthmatic children thanks to a serious game (Delmas, et al., 2018). Hence, activity selection might be crucial to elicit and maintain students' engagement in the task and would in turn yield enhanced learning. Such results are in accordance with the developmental hypothesis of learning progress (Oudeyer, et al., 2016) by which the learner is intrinsically motivated to practice tasks that elicit optimal learning gains. Learning gains are thus seen as an internal motivational reinforcer for learners. By providing personalized learning paths, the ZPDES maximizes each learner's progress and stimulates motivation and perseverance in learning (Metcalfe & Kornell, 2005).

Importantly, ZPDES provides a full adaptability of the instruction curriculum for each student, and then, is able to respond to the learners' diversity. The ZPDES-related adaptability for interindividual differences in learning mirrors the purpose of disability-related instructional design frameworks (*e.g.*, Universal design for learning, Rose & Meyer, 2006; Goodall, 2015), in which the ultimate goal is to provide common principles for successful intervention for all, while addressing idiosyncrasies of each one. The association between motivation and learning is documented as a critical factor in fostering learning in individuals with ASD (Keen, 2009; Koegel, et al., 2010). Several studies leveraged the sensitivity of individuals with ASD to positive reinforcers for developing their adaptive behaviors (*e.g.*, Schreibman, et al., 2015; Wong, et al., 2015). Furthermore, some ASD studies reported engagement difficulties in school-related tasks when they are too easy or too difficult for the individual with ASD (Koegel, et al., 2010). At the best of our knowledge, no ITS-based pilot study with ASD children has been conducted with the characteristic of proposing a personalization approach based on learning progress and its optimization in the activities space as a guide for addressing the zone of proximal development.

Aim of the study

Using ITS with students in specialized classrooms can be a promising approach to provide dynamically individualized educational interventions. To explore this research avenue, we conducted a pilot study with secondary-school students with ASD and/or ID included in specialized classrooms, in order to examine the feasibility and the relevance of an ITS-based intervention for instructing basic mathematical abilities (*i.e.,* adding and subtracting numbers from 0 to 100) through monetary exchanges activities.

Our primary purpose is not to evaluate the added value of an ITS based on learning progress compared to another pedagogical condition, but rather to evaluate the relevance and possible interest of such an intervention in order to later, if conclusive, conduct a comparative study to estimate its added value pedagogically. Simply formulated, our primary purpose is to explore whether children with ASD can be responsive to this type of intervention or not. To this end, we used the *KidLearn* tablet-based application including the ZPDES algorithm (Clément, et al., 2013). The addressed research questions were:

- To what extent *KidLearn* intervention may improve addition and subtraction performances of students with ASD and/or ID?
- To what extent *KidLearn* intervention may foster the students' learning progress and motivation?

Materials and Method

As a result of our exploratory purpose, this study employed a quasi-experimental pre-post design with an active control condition. This research design consisted in comparing two independent groups receiving two distinct interventions, with pre-post assessments. Then, the children received either the ITS-based educational instruction (experimental condition) or an ASD-specific instruction targeting gaze orientation and face processing (control condition), both during group sessions led by the experimenter. Using active control condition instead of no-treatment or treatment-as-usual condition is employed in interventional trials in order to, for example: control the effect related to attention received by the research staff, control nonspecific treatment effects, and prevent control group attrition (Lindquist, et al., 2007). This kind of design is commonly used in interventional trials for children with ASD in natural settings, including schools (*e.g.*, Kenworthy, et al., 2014; Soorya, et al., 2015; Spaniol, et al., 2018). Additionally, we included measures related to the potential outcomes of the control intervention to monitor its specific effects.

The protocol and the consent form were examined and approved by the ethical committee of authors' research center (COERLE[1]) and CNIL[2] before the beginning of the study. The parents' and students' consents were collected after they were informed of the conditions of the study.

Participants

Recruitment

Our study took place in specialized classrooms of several secondary schools, in the context of a research-action project. We initially recruited 33 students aged from 11 to 16 years-old across four secondary schools located in the south-west territories of France. The recruitment of participants was supported by the academy rectorship for contacting schools with specialized classrooms. Then, we met the headmaster and the specialized teacher of volunteer schools to assess their eligibility (having students with ASD and/or ID in specialized classrooms). French specialized classrooms receive students with various disabilities who benefit from an individualized instruction. These classrooms are also called inclusive in France because students may be included from few hours to few days per week in regular classrooms.

Participants were 6th-to-9th grade students with cognitive disabilities enrolled in inclusive specialized classrooms, having either ASD, ID, or both. Participants were allocated in two matched groups: one group received the *KidLearn* tablet-based intervention, while the other received another tablet application as a control intervention. This setting allowed us to control the contextual bias of providing a tablet-based intervention to one student. As the application for the control group did not include any mathematical content, we hypothesized that only the group equipped with *KidLearn* will increase their mathematical abilities. Our protocol also included supplementary measures for visual general and face processing, in order to control for external validity.

Group matching procedure

A first group matching procedure was carried out based on children's age and SRS score. Before the beginning of the study, specialized classrooms' teachers filled out the SRS-II in order to evaluate the level of ASD severity in each recruited student. The SRS-II (Social Responsiveness Scale, Constantino & Gruber, 2012) is a parent- or teacher-reported questionnaire which is often employed as a screening tool for ASD (Schanding, et al., 2012), as this test may capture the severity of ASD symptoms (Naglieri & Chambers, 2009; Schanding, et al., 2012; Stordeur, et al., 2019). Then, the SRS was used as a matching criterion for balancing ASD severity in the two groups. The SRS is a 65-item questionnaire which assess the frequency of ASD-related behaviors (*e.g.*, communication and social

---

[1] COERLE stands for "operational committee for the assessment of legal and ethical risks". This committee is responsible for examining research projects and experimental studies in terms of legal and ethical issues. The COERLE is seized before the study to verify the risks of harm to individuals.

[2] CNIL stands for "national committee of informatics and liberty". The CNIL in a French independent administrative authority responsible for ensuring that technological products and services do not infringe citizen's identity and privacy, and that they respect human rights and individual freedom.

interactions, repetitive and restricted behaviors). The total score ranges from 0 to 195 and reflects the symptom severity: the higher the score, the more severe ASD symptoms. The SRS has good to excellent psychometric properties (Naglieri & Chambers, 2009). The French version of the SRS showed a good discriminant validity for detecting children with ASD (Stordeur, et al., 2019), based on a ROC analysis ($0.70 < AUC < 0.75$).

The matching procedure was carried out by the experimenter: the two groups were manually balanced in order to have comparable SRS score and age means. The quality of the matching was checked using effect size estimates (Cohen's *d*) and variance ratios, as advised in Kover & Atwood (2013). The matching procedure was controlled with a mean comparison test: for age, $t(31) = 0.021$; $p > .900$; for SRS: $t(31) = 0.083$; $p > .900$, and resulted in two groups of 16 and 17 students.

Inclusion/Exclusion criteria

The attendance at all evaluation and intervention sessions was required for participants to be included in the study. Seven out of 33 initially recruited students were excluded because of their absenteeism during evaluation and/or intervention sessions. Another inclusion criterion was the intellectual functioning, as assessed by an abbreviated version of the WISC-IV (*i.e.*, Similarities, Symbols, Letter-Number Sequences, Matrix Reasoning; Grégoire, 2009; Wechsler, 2003). This abbreviated version was designed on the basis of the most reliable and correlated subtests and provide a reliable estimation of the full IQ (reliability coefficient: 0.91). The assessment of IQ resulted in the exclusion of two additional students (*i.e.*, $IQ > 90$).

The final sample was composed of 24 students, divided between the intervention and the control groups. The intervention group was composed of 14 students (5F;9M) and received the *KidLearn* tablet application, equipped with the ZPDES algorithm. The control group was composed of 10 students (4F;6M) and received the "control" tablet application. Gender distribution did not differ between the groups (Fisher's test: $F = 1.19$; $p > .900$).

The quality of matching was controlled afterwards between the final groups (Table 1): the two groups did not significantly differ in age ($Z = -0.149$; $p > .800$; $d = 0.11$), either on the SRS score ($Z = 0.937$; $p > .300$; $d = 0.25$). The raw SRS scores ranged from 10 to 156 in the control group (M = 65.10), with 6/10 students above the ASD clinical threshold. In the intervention group, raw SRS scores ranged from 12 to 112 ($M = 54.43$), with 6/14 students above the ASD clinical threshold. The IQ of participants ranged from 35 to 86 in both groups, with an average of 55.14 in the intervention group and 62.70 in the control group. Between-group differences in IQ were not statistically significant ($Z = -1.202$; $p > .200$; $d = 0.51$).

A profile questionnaire (see *Materials* section) gave additional information on the students' appetence for calculation and the screens' frequency of use. The students' repartition on these variables did not significantly differ between the groups (Fisher's test: $p > .500$). The majority of participants stated that they like performing calculation and that they frequently use screens (*i.e.*, 19/24 students use digital devices more than 4 days a week).

Table 1. *Descriptive statistics at pre-intervention*

	Intervention (N=14)		Control (N=10)		*Z*	*p*	*d*
	M	*SD*	*M*	*SD*			
Age	13.36	1.69	13.20	1.40	-0.149	.881	0.11
IQ	55.14	14.43	62.70	16.17	-1.203	.229	0.51
SRS	54.43	44.99	65.10	38.58	-0.937	.349	0.25
Pre-test							
Calculation test	0.35	0.26	0.38	0.18	-0.265	.791	0.14
Transfer questions	0.45	0.43	0.47	0.28	-0.122	.903	0.00
NEPSY	18.89	5.50	20.00	3.17	-0.029	.977	0.25
KABC-Seq	21.79	5.38	20.80	3.88	-0.471	.638	0.21
KABC-Sim	34.36	12.19	42.00	13.97	-1.085	.278	0.61

Materials

KidLearn Application

The *KidLearn* application was designed to instruct basic calculation and money manipulation through monetary exchanges activities. Two types of exercise were proposed, in which the student may be the customer or the merchant. Then, the student either had to compose the sum to buy one or two items, or to give change to a customer who buys one or two items (Figure 1). Both types of exercises were proposed to the children. The ZPDES algorithm managed the type of exercises provided to the children, as the "merchant" exercises were considered more difficult than the "customer" exercises (*i.e.*, subtracting vs. adding). Consequently, students may complete merchant and/or customer exercises in each session depending on their progress.

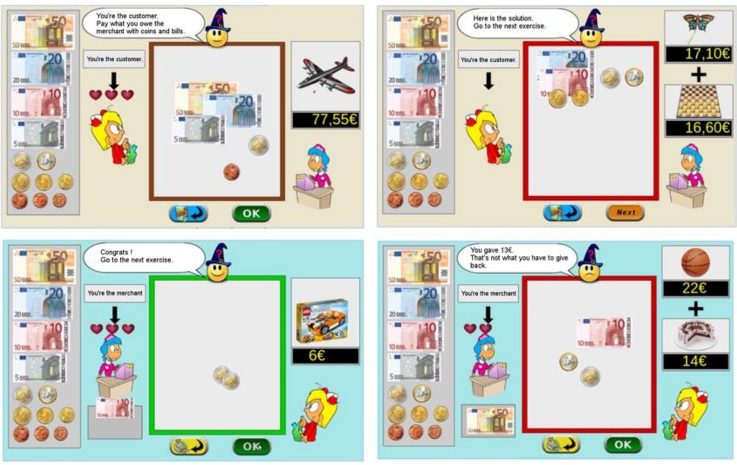

Figure 1. *Visuals of the KidLearn application. Taken and translated from Clément (2018). (a) – Top, left*: the student plays the customer role and must place the banknotes and coins needed to purchase an item. *(b) – Top, right*: the student plays the customer role and must purchase two items. Here, the feedback provides the answer after three unsuccessful attempts. *(c) – Bottom, left*: the student plays the seller role and successfully gives changes for the purchase of an item. *(d) – Bottom, right*: the student plays the seller role and failed to give change for the purchase of two items. Here, the feedback indicates that he does not give the correct amount of money.

At the beginning of the activity, one or two items with their corresponding prices were displayed on the screen. The student had to drag and drop the money they wanted to use from the wallet location to the receiving location. When the student needed help, they could click on the character in the top of the screen to get a clue. Once the student had finished, they had to click on the "OK" button to confirm his/her answer. The student then received feedback on the response. If they succeeded, they were congratulated and moved on to the next exercise. If the answer was incorrect, a clue was provided (for example: *"You placed X €, this is not what you have to give to the customer"*) and the student could try again to find the answer. After the third consecutive unsuccessful attempt, the system would feedback the correct response and then switch to another exercise.

The *KidLearn* application was initially developed for instructing elementary-level neurotypical children, and the design of the user interface was not built according to ASD-specific guidelines. There is no consensus on ASD-specific guidelines for designing user interface in tablet-based app, but some studies provided some recommendations about user interface for individuals with ASD (Dattolo & Luccio, 2017a, 2017b; Grynszpan, et al., 2008; Pavlov, 2014). The screening of the *KidLearn* interface showed that it is globally suited for students with ASD and that there were no major accessibility issues. Indeed, the interface structure is simple, clear and predictable: every element is clearly identified and clustered on the interface (*i.e.*, money wallet on the left of the screen, receiving area at the center, the object to be bought on the right, controls buttons on the bottom). Feedbacks are provided through visual and textual outputs (*i.e.*, receiving area's outline colored in green or red, and textual feedback from the character). Also, customer vs. merchant exercises were distinguished through background coloring, the placement of customer and merchant characters, and textual information. Using photographs of real objects and money is also a good feature for students with ASD, as it refers to concrete objects from the environment, and may support the skill transfer. The navigation is simple, as the access to the game is driven by a separate teacher interface, and the child has only two controls for either validating or erasing the answer that they compose in the receiving location.

The difficulty level was determined by two parameters: 1) the number decomposition difficulty and 2) the use of cents. A number is easy to decompose when it is in direct relation with real banknotes and coins (*i.e.*, 0, 1, 2, 5), whereas a number is harder to decompose when it requires to associate several banknotes and/or coins (*i.e.*, 3, 4, 6, 7, 8, 9). The number decomposition difficulty is also related to the use of carries: the computation is harder when they require a carry (*e.g.*, 15-6, 9+4), than when there is no carry (*e.g.*, 15 – 2, 6 + 3). The use of cents implies the manipulation of decimals rather than integers, which increases the complexity level of calculation.

ZPDES algorithm (Clément, 2018; Clément, et al., 2013). From the complete activity sequence, the ZPDES algorithm is initialized with a subset of initial activities that can be proposed to the student. This activity space is readjusted depending on the students' performance by adding or deleting activities. This activity space is called "zone of proximal development" (ZPD), inspired from the Vygotsky's concept (Vygotsky, 1980).

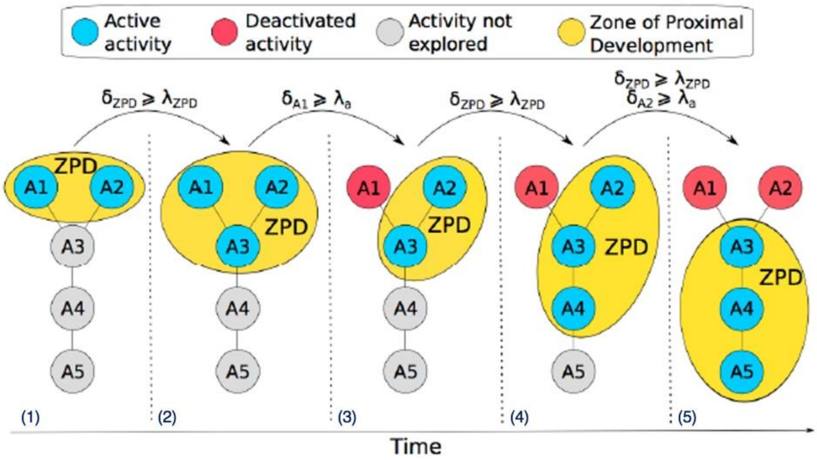

Figure 2. *Operating diagram of the ZPDES algorithm. Taken from Clément (2018). From left to right*: The ZPDES algorithm first initializes a ZPD space with a set of starting activities (1). The A1 and A2 activities are activated and may be proposed to the learner. When the success rate on the ZPD space exceeds a fixed threshold ($\delta_{ZPD} \geq \lambda_{ZPD}$), a new activity is added to the ZPD space (2). When the success rate on one activity exceeds a fixed threshold ($\delta_{Ai} \geq \lambda_A$), this activity is deactivated and then removed from the ZPD space (3). This process is repeated across the trials (3, 4, and 5) and allows to generate learning spaces adapted to each student's progress.

Figure 2 depicts the global functioning of the algorithm. The ZPD is first initialized with a set of activated activities that can be proposed to the learner. When the success rate in the ZPD space (δ_{ZPD}) exceeds the activation threshold (λ_{ZPD}), the ZPD is extended to include further activities. When the success rate on a single activity (δ_{Ai}) exceeds the inactivation threshold (λ_A), this activity is deactivated from the ZPD and will not be proposed to the student anymore. In this way, the ZPDES algorithm is able to propose activities from a custom set of "learnable" activities that are neither too easy nor too difficult for a given student. This algorithm may then maximize the learning gain and foster the learners' motivation.

"Control" Application

The "control" application consisted in presenting a picture to the student, which depicts a person or a social scene (Figure 3). For each picture, a set of symbols has been defined to be briefly depicted on up to three points of social interest in the picture. Once the symbols disappeared, the student had to recall the nature and the number of depicted symbols.

Three difficulty levels were defined based on the content of the pictures, with: 1) pictures of faces, portraits with one person; 2) simple social scenes, with two or three individuals and, 3) complex social scenes with more than three individuals. The set of pictures was selected from the socio-moral picture database, composed with pictures either in the public domain, or under open licenses (Crone, et al., 2018). Another difficulty parameter was determined by the duration for which the symbols appeared. For the three difficulty levels, the speed of appearance was set depending on three modalities: slow (2s), moderate (1s) and fast (0.2s). In the "fast" modality, tracking the symbols requires to keep focused attention on the picture and its content. Before the symbols appear, the picture was displayed for 3 seconds. The response keyboard was displayed 5s after the symbols disappeared.

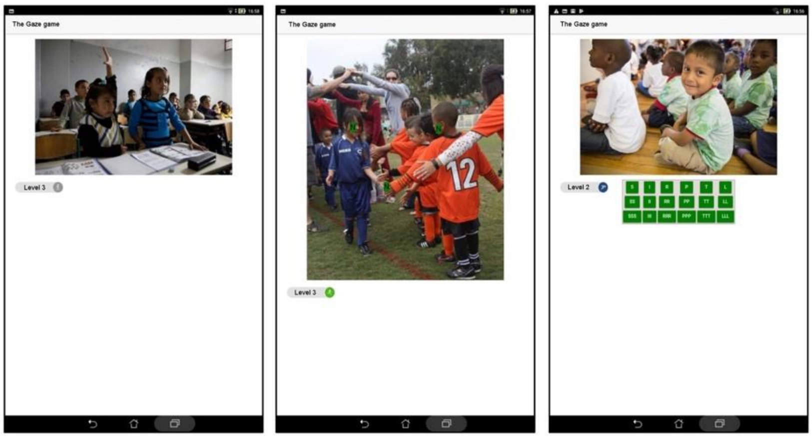

Figure 3. *Visuals of the control application (gaze orientation training). (a) – Left*: A picture is displayed on the screen. The student must focus his/her attention on the important individuals in the picture. Below the picture, we can see the actual level at which the student is (i.e., level 3), and the level speed (i.e., grey walking character next to the level indication; here, the level is low speed). *(b) – Middle*: One to three symbols appear on the points of interest of the picture, and the student must identify the nature and the number of symbols. Below the picture, information on the present level indicates that the student is at level 3 with a moderate speed level (i.e., green jogging character). *(c) – Right*: Once the symbols disappear, a response keyboard appear below the target picture. The student must click on the right answer (i.e., which symbol and how much). Below the picture, we can see that the student is actually at level 2, with a fast speed level (i.e., blue fast running character).

The student moved from one level to another depending on their success and failure rates: a success rate of 70% was required to move on to the next level, while a failure rate of 20% resulted in moving back to the previous level. In order to ensure a minimal amount of instruction opportunities at each level of the application, the student had to perform a minimum of 50 trials before being able to pass to another level.

Dependent variables

Profile questionnaire. A profile questionnaire has been developed to screen general information on each student: *e.g.*, gender, screen usage frequency, perceptions and habits regarding the use of money. The English translation of the profile questionnaire is provided as a supplementary material.

Internal measures.

Calculation test. To consider the internal validity of the intervention's effects, we evaluated calculation performance of the students using a 20-item calculation test, composed with: one question on money knowledge, in which the student had to select the existing banknotes and coins from those proposed; three calculation questions implying flowers bouquets (transfer questions) in which the student has to compose flower bouquets using packages of 20, 10, 5, 2 and 1 flower(s) (the number of flowers needed had to be calculated from the information included in the statement); and 16 calculation questions on monetary material, with variable difficulty (*i.e.*, buying one item with and without cents; buying two items with and without cents; giving change on the purchase of on item, with and without cents; giving the change on the purchase of two items with and without cents). The total scores were computed with the correct proportion of correct responses.

Two similar versions of the calculation tests have been elaborated in which only the numbers are modified for performing the calculation: mathematical operations were the same and the numbers are selected in quasi-random manner among a pre-selected range of numbers. This allowed us to provide similar versions of the calculation test with a similar difficulty level.

Performance monitoring. To monitor the children's progress, we computed a score at five timepoints across the whole intervention (*i.e.*, 10^{th}, 25^{th}, 50^{th}, 75^{th} and 100^{th} exercise). This score was computed based on the number of exercises successfully resolved in each exercise group, with the following formula: $1 \times N_{Grp1} + 2 \times N_{Grp2} + 3 \times N_{Grp3} + 4 \times N_{Grp4}$. The groups refer to the four exercise groups, increasing in difficulty of calculation: 1) exercises as the customer with one item (one addition); 2) exercises as the merchant with one item (one subtraction); 3) exercises as the customer and two items (two additions), and 4) exercises as the merchant and two items (one addition and one subtraction).

Motivation. Motivational levels of the student were evaluated using the motivation questionnaire of Vallerand, *et al.* (1989). This 21-item questionnaire assessed the level and the type of the learners' motivation at the end of the intervention. The content of the questionnaire was adapted to our setting of a children's game. Each item stated a reason for being motivated to play or not: the student must answer to each question with "yes" or "no". The total score (max. 21) reflects the student's global level of motivation: the higher the score, the higher the motivation is. Items are spread across three dimensions, corresponding to the different kinds of motivation described in the self-determination theory (Ryan & Deci, 2000). This questionnaire allowed us to quantify the level of motivation of each type: 1) amotivation (/3) is related to the absence of motivation for an activity: individuals do not relate their actions to the results they obtained and perceive the situation as the result of external factors (example: "I played the game but I don't know why", "I played the game because I did what I was told to do"); 2) intrinsic motivation (/9) is related to carrying out an action for one's own satisfaction, for instance, an intrinsically motivated individual carries out an action for the sense of challenge or for the pleasure, but not for an external reward (example: "I played the game because I like to learn new things", "I played the game because I like to do well at a game"); and 3) extrinsic motivation (/9) is related to carrying out an action for reaching an external aim, such as a reward (example: "I played the game to be allowed to use the tablet", "I played the game because I would have been ashamed if I hadn't try").

External measures.

For considering the external validity of the intervention's effects, we have chosen measures related to the use of tablets (*i.e.*, general visual processing) and to the specific use of the control application (*i.e.*, specific visual processing of faces). Indeed, the use of tablets may affect the general visual processing in children, and the use of an app with social pictures may affect related socio-cognitive processes.

General visual processing. We selected four subtests from the Kaufman Assessment Battery for Children (KABC-II, Kaufman & Kaufman, 2004).

- *Sequential processing* (*Gsm*), with a max score of 50, and composed of two subtests: 1) *Number recall* for which the child had to repeat a series of number sequences in the same order. The number sequences were long from 2 to 9 items, with a progressive increase across the trials. The maximal score for this subtest was 22; 2) *Word Order* for which the child had to recall a series of spoken words in the same order, using a pictures' sheet representing the words. This test was performed in two parts, with series from 2 to 6 words. The first part implied an immediate recall of the words' series, and the second implied a color-reading interference task before performing the recall. The maximal score for this subtest was 28.

- *Simultaneous processing* (*Gv*), with a max score of 79, and composed of two subtests: 1) *Rover* for which the child had to move a toy dog on several grids of variable dimensions (from 4x4 to 6x6) and to find the quickest path to find the bone. The child also had to respect two constraints: a) some cases were forbidden (*i.e.*, brambles), and b) some cases counted for two moves instead of one (*i.e.*, rocks). This task was timed and had to be completed within a limited time. The maximal score for this subtest was 38, depending on the answer and the time to complete each trial. 2) *Triangles* for which the child had to reproduce figure models using two-colored foam triangles. The test started with simple figures composed of two triangles and gradually increased in complexity until 9 triangles were used. The child had to complete each trial in a limited time. The maximal score for this subtest was 41, depending on the answer and the time to complete each trial.

The KABC-II assessment is standardized and has good psychometric properties (Kaufman, 2005; Kaufman & Kaufman, 2004). The test-retest reliability[3] coefficient lied between .74 and .78 for the simultaneous processing score (*Gv*), and comprised between .79 and .80 for the sequential processing score (*Gsm*).

Visual face processing. We selected two subtests of the Developmental Neuropsychological Assessment (NEPSY-II, Korkman, et al., 2007): 1) *Memory for Faces* which started with a learning phase in which 16 faces were sequentially presented to the child; the encoding is encouraged by asking the child to tell whether the face is a girl or a boy. Once the learning phase was completed, an immediate recall was performed using 16 sheets, each depicting three faces (including the target and two distractors). For each trial, the child had to select the picture he had learned during the learning phase. A delayed recall was performed 15 to 25 minutes after the immediate recall, using 16 sheets depicting three faces as well. The total score for this subtest was 32 (the sum of correct answer in the immediate and the differed

[3] The test-retest reliability is a psychometric quality of great importance in pre-post assessments, since it considers the stability of the scores of the same individual over time. The test-retest reliability is assessed through the examination of the stability of the scores between two assessments, using intra-class correlations. The higher the correlation, the more stable the score are over time. If the coefficient is below .70, the test is likely to be sensitive to learning effects (Bernaud, 2014).

recall. 2) *Affect Recognition* consisted in showing a series of faces to the child and asking him to associate faces that depicted the same emotion. Several conditions were proposed, with for example, linking of two faces or finding the face which fits with a target one. The maximal score for this subtest was 35. The scores on these two subtests have been averaged to compute a unique measure related to the visual processing of faces, with a maximal score of 33,5. These assessments are standardized and have good psychometric properties (Brooks, et al., 2009; Korkman, et al., 2007). For the *Affect Recognition* subtest, the test-retest reliability coefficient was .55 for the 11-12 age group, and .58 for the 13-16 age group. For the *Memory for Faces* subtest, the test-retest reliability coefficient was .57 for the 11-12 age group, and .73 for the 13-16 age group.

Procedure

The study was performed between March and June 2019, in four specialized classrooms. The assessment of the tablet applications was performed using a pre-post comparison, consisting in assessing the outcomes of interest before and after the intervention (Figure 4). All children enrolled in this study continued to attend their regular classes, including math classes, except when the time slot was devoted to an intervention session.

Pre- and post-interventions evaluations were carried out through individual interviews between the experimenter and each participating student. These interviews took place in a school room provided for this purpose, for less than one hour at a time. Two to three interviews per student were conducted to perform the full evaluation and avoid fatigability issues at pre- and post-interventions assessments. Only the calculation test was carried out during the intervention period because the test was integrated in the *KidLearn* interface. The pre-test was carried out at the beginning of the first intervention session, and the fourth session was dedicated to the post-test (there was no instruction during this fourth session, but the students could freely play to the game after the assessment if they were willing to). During the calculation test, the use of a calculator was not authorized, but the student could use their blackboard or a draft notebook for performing the calculation if they needed. After the fourth session was completed, post-intervention assessments were conducted using the same procedure as the pre-intervention assessment.

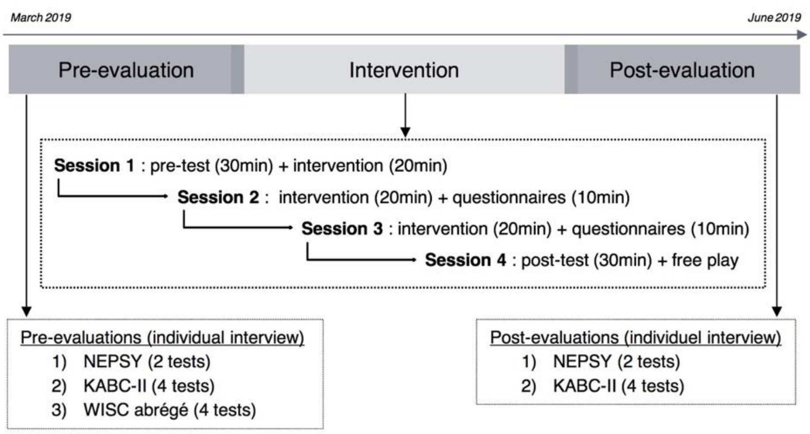

Figure 4. *Graphical representation of the process of the intervention*

The intervention sessions started once individual evaluations of all the students were carried out. The sessions took place in the specialized classrooms or in a room provided for this purpose, and each session lasted half an hour, with 20 minutes of tablet activity. The experimenter brought the materials and animated the sessions in each classroom. The number of students in each session was comprised between 2 and 10 students, and each participant followed all the sessions. Each student received a tablet for using the application intended for them. Four intervention sessions were followed by the participants (Figure 4), with three instruction sessions and a fourth session dedicated to the post-test (integrated in the *KidLearn* application). For the second and third sessions, the children picked up the exercises at the level they left off at the previous session.

The experimenter, the specialized teacher and the school aides assisted some students by providing counting strategies for facilitating the calculations. They helped students either when they were stuck and couldn't do no exercises at all, or when they requested it because they were not able to master the activity. The help provided did not consist in giving the solution, but in suggesting strategies for solving the calculation (*e.g.*, providing concrete manipulatives, helping in decomposing the numbers, helping in finding errors).

Finally, as this study aimed at examining the feasibility and the relevance of using KidLearn with students having ASD and/or ID, the experimenter observed the course of the sessions and noted the points of blockage or difficulty related to the use of an ITS in a specialized classroom.

Data analyses

The data analyses were performed using SPSS version 22. Mean comparison tests were performed on internal and external measures for examining pre-post effects and between-group differences. Pre-post differences were assessed using a mean comparison test for paired samples, and between-group differences with a mean comparison test for independent samples. As the prerequisites for parametric testing were violated, we selected their non-parametric counterparts (*i.e.*, Mann-Whitney test for independent samples, Wilcoxon test for paired samples). Even if parametric analysis is relatively robust to normality assumption violations, non-parametric tests have more power than their parametric counterparts, when there are both unequal small sample size ($N < 30$ per group) and non-normal data distribution (Pett, 2015). In several cases, when normality assumptions were violated, the data distributions were either positively or negatively skewed. Given the small and unequal sample size and the distribution skewness, we carried out non-parametric analyses as they have more power than the classical parametric procedure in this specific context (Vickers, 2005).

Finally, post-hoc analyses were carried out by examining the contribution of intellectual efficiency and autism severity to the students’ performance. For this purpose, we computed a correlation matrix for each group with internal and external measures, as well as IQ estimates and raw SRS score.

Results

The Table 1 depicts pre-intervention measures for controlling the quality of group matching. Means, standard deviations and pre-post assessments for internal and external measures are depicted in Table 2. The Table 3 provides performance monitoring scores computed in the intervention group. Motivation scores and their between-group comparisons are depicted in Table 4.

Table 2. *Pre-post effects of intervention*

	Pre-test		**Post-test**				
Intervention group (N = 14)	*M*	*SD*	*M*	*SD*	Z	*p*	*d*
Calculation test	0.35	0.26	0.46	0.26	-2.010	.044	1.34
Transfer questions	0.45	0.43	0.48	0.43	-0.172	.836	0.09
NEPSY	18.89	5.50	20.60	4.93	-2.630	.009	2.13
KABC-Seq	21.79	5.38	21.50	4.35	0.276	.782	0.18
KABC-Sim	34.36	12.19	39.21	14.97	-2.477	.013	1.64
Control group (N = 10)	*M*	*SD*	*M*	*SD*	Z	*p*	*d*
Calculation test	0.38	0.18	0.42	0.27	-0.655	.512	0.46
Transfer questions	0.47	0.28	0.43	0.42	-0.333	.739	0.22
NEPSY	20.00	3.17	18.00	5.44	-1.735	.083	1.34
KABC-Seq	20.80	3.88	21.70	5.14	-0.632	.527	0.49
KABC-Sim	42.00	13.97	48.10	14.85	-2.565	.010	2.28

Table 3. *Performance monitoring score at each timepoint in the intervention group.*

10th exercise		**25th exercise**		**50th exercise**		**75th exercise**		**100th exercise**	
M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
3.00	1.96	8.03	4.23	14.66	6.41	16.23	8.21	15.83	7.67

Table 4. *Motivational measures and between-group comparison at post-intervention*

	Intervention (N=14)		**Control (N=10)**		*Z*	*p*	*d*
	M	*SD*	*M*	*SD*			
Motivation score	14.64	5.08	4.90	5.95	-3.092	.002	1.84
Amotivation	2.07	0.83	0.70	1.06	-2.900	.004	1.52
Intrinsic motivation	6.50	2.28	2.30	2.75	-3.137	.002	1.64
Extrinsic motivation	6.07	2.84	1.80	2.44	-3.133	.002	1.74

Between-group Comparison at pre-intervention

Internal measures.

The calculation score in the intervention group did not significantly differ from the control group on pre-intervention data (Z = -0.265; $p > .700$; d = 0.14). Both groups displayed a similar calculation performance before the intervention ($M_{Intervention}$ = 35% vs. $M_{Control}$ = 38%).

External measures.

On the score derived from the two NEPSY subtests, no significant between-group differences were evidenced (Z = -0.029; $p > .900$; d = 0.25). We obtained similar results for the two KABC-II scores, with no significant between-group differences neither on sequential processing score (Z = -0.471; $p > .600$; d = 0.21), nor on simultaneous one (Z = -1.085; $p > .200$; d = 0.61). In summary, the two groups had similar performances on all measures before the intervention began.

Within-group Comparison: pre-post effects.

Internal measures.

The pre-post effect was significant for the calculation test in the intervention group (Z = -2.010; $p < .05$; d = 1.34), but not in the control group (Z = -0.655; $p > .500$; d = 0.46). The calculation performance of the students in the intervention group on the post-test (M_{POST} = 46%) was significantly higher than performance on the pre-test (M_{PRE} = 35%). In contrast, the performance of the students in the control group on the post-test (M_{POST} = 42%) did not significantly differ from their performance on the pre-test (M_{PRE} = 38%).

The two groups were compared on the level of motivation at the end of the intervention. The level of motivation elicited by the application was significantly different between the two groups (Z = 3.092; $p < .01$; d = 1.84). The total motivation score was much higher in the intervention group ($M_{Intervention}$ = 14.64) than in the control group ($M_{Control}$ = 4.90). This difference was also observed on the sub-dimensions of the motivation test: 1) amotivation (Z = 2.900; $p < .01$; d = 1.52), 2) intrinsic motivation (Z = -3.137; $p < .01$; d = 1.64), and 3) extrinsic motivation (Z = -3.133; $p < .01$; d = 1.74). Therefore, the *KidLearn* application elicited a higher motivational level than the control application, irrespective of the type of motivation.

For further information about the children's progress in the intervention group, a score for performance monitoring was computed based on exercises groups successfully reached by each child (Table 3). In average, we observed that the performance monitoring score improve with the time ($F(1,13) = 33,62$; $p < .001$). Paired comparisons with Bonferroni corrections showed significant differences between 10^{th} and all other timepoints ($p < .01$) and between 25^{th} and all other timepoints ($p < .01$). The difference between 50^{th} and 75^{th} exercise timepoint was close to the .05 threshold ($p = .065$). The difference between 50^{th} and 100^{th} exercise, and between 75^{th} and 100^{th} exercises were not significant. From these results, we may infer that the children in the intervention group exhibited improvement in calculation, until they reached the 50^{th} exercise, and then, maintained this performance.

External measures.

The pre-post effect on the NEPSY score was significant in the intervention group (Z = -2.630; p < .01; d = 2.13), but not in the control group (Z = -1.735; p > .05; d = 1.34). The students in the intervention group had a post-intervention score (M_{POST} = 20.61) significantly higher than the score at pre-intervention (M_{PRE} = 18.89). By contrast, the difference between pre- and post-score in the control group was not sufficient to be statistically different (M_{PRE} = 20.00 and M_{POST} = 18.00).

On KABC-II measures, a significant pre-post effect was evidenced for the simultaneous processing score in the intervention group (Z = -2.477; p < .02; d = 1.64), as well as in the control group (Z = -2.565; p < .01; d = 2.28). Both groups had a post-intervention score significantly higher than the pre-intervention score (Table 2). However, no significant pre-post effect was evidenced on the sequential processing score, neither for the intervention group (Z = 0.276; p > .700; d = 0.18), nor for the control group (Z = -0.632; p > .400; d = 0.49).

Post-hoc analyses: contributions of IQ and autism severity to the intervention results.

Table 5 depicts the correlation matrix for examining the influence of IQ and autism severity (SRS raw scores) on the students' performance in both groups. Significance threshold have been corrected using the Bonferroni method.

In the intervention group, we found significant correlations between the SRS raw score and the NEPSY score at pre- (r = -.80) and at post-intervention (r = -.83). These strong and negative correlations suggested that the increase in the NEPSY score observed in the intervention group, may be negatively mediated by the autism severity measure. The students with a low autism severity were those who have best succeeded in the NEPSY subtests at pre- and post-intervention.

Table 5. *Correlation matrix of IQ and SRS raw score with outcome measures.*

	Intervention group				**Control group**			
Variables	*N*	*M*	*1*	*2*	*N*	*M*	*1*	*2*
1. IQ	14	55.14	-		10	62.70	-	
2. SRS	14	54.43	-.31	-	10	65.10	-.38	-
3. NEPSY Pre-test	14	18.89	.61	**-.83***	10	20.00	.69	-.46
4. NEPSY Post-test	14	20.60	.42	**-.80***	10	18.00	.70	-.41
5. KABC-Seq Pre-test	14	21.79	.44	-.32	10	20.80	.53	-.20
6. KABC-Seq Post-test	14	21.50	.71	-.24	10	21.70	.86	-.58
7. KABC-Sim Pre-test	14	34.36	.68	-.48	10	42.00	.87	-.32
8. KABC-Sim Post-test	14	39.21	.72	-.61	10	48.10	.71	-.19
9. Calculation Pre-test	14	0.35	.36	-.18	10	0.38	.65	.01
10. Calculation Post-test	14	0.46	.45	.03	10	0.42	.71	.13

Significant correlations (p < .05 with Bonferroni's correction)

The performance of the students in the intervention group on the other measures was not significantly correlated neither with the SRS raw score, nor with the IQ. In the control group, we did not find any significant correlation between the students' performances and the IQ or the autism severity measure.

Discussion

This pilot study aimed at examining the feasibility of an educational intervention based on an ITS for instructing mathematical abilities in secondary school students enrolled in specialized classrooms. Precisely, we examined the efficacy of the ZPDES algorithm in reinforcing the linkage between learning gains and motivation in secondary students with ASD, low mathematical abilities and a moderate-to-severe ID.

Mathematical and motivational outcomes

First, the results supported the relevance of the *KidLearn* application equipped with the ZPDES algorithm for secondary students with ASD and/or ID. Indeed, *KidLearn* application yielded positive effects on the calculation performance. The children in *KidLearn* condition achieved an average of about 9.1/20 on the calculation post-test, with a gain of about 2 points in just three sessions of use. Also, the performance monitoring score showed that the children in the KidLearn condition constantly improve their performance to the 50th exercise. By contrast, the students in the control group gained less than one point on the calculation test after the intervention ($M_{PRE} = 7.70$ and $M_{POST} = 8.50$). Concomitantly, learning progress in calculation performance was associated with a higher motivational level in children in the *KidLearn* condition than those in the control condition. Taken together, the results are consistent with the developmental hypothesis of learning gains (Keen, 2009; Oudeyer, et al., 2016), in which the learning-motivation linkage is emphasized.

The results of this study are also consistent with those obtained by Clément (2018), which evaluated the *KidLearn* application and the ZPDES algorithm with typically-developing primary school pupils (7-8 years old). Their participants had a starting score on the calculation test between 5.9 and 7.7, and they increased on average by three points at the end of the intervention. The level of motivation of their participants (*i.e.*, intrinsic, extrinsic, global) was also comparable to those of our sample. The fact that the students of our sample had a calculation level comparable to that of primary school pupils was not surprising given the literature on the mathematical abilities of students with ASD (Bullen, et al., 2020; Chiang & Lin, 2007; Fleury, et al., 2014; Jones, et al., 2009; Morsanyi, et al., 2018; Oswald, et al., 2016; Wei, et al., 2015) and those enrolled in specialized classrooms (Fleury, et al., 2014; Jones, et al., 2009; Kurth & Mastergeorge, 2010; Keen, et al., 2016). The *KidLearn* application addressed basic mathematical abilities, which were identical to those successfully addressed by the interventions targeting secondary students with ASD and/or ID (*e.g.*, Bouck, et al., 2020; Cihak & Grim, 2008; Fletcher, et al., 2010; Hsu, et al., 2014; Spooner, et al., 2019). Correlational analyses did not show a mediating effect of IQ or SRS raw score on the calculation test results. The increase in calculation scores observed in our sample cannot be therefore related to the level of intellectual efficiency or the presence and the severity of ASD.

Another question we considered, was whether the achieved calculation progress with *KidLearn* is generalizable to domains other than money. To this end, we included a transfer task based on counting flowers. We did not observe any significant progress on the three transfer items proposed in the calculation test. This result fits with studies reporting generalization difficulties in individuals with ASD (*e.g.*, Fleury, et al., 2014; Spooner, et al., 2019). Additionally, it would be interesting to explore more deeply the generalization of acquired skills using a more ecological task implying a real purchasing situation, as it was done in some other interventional studies (*e.g.*, Cihak & Grim, 2008; Fletcher, et al., 2010).

Findings on external measures

The pre-post effects revealed an increase in the NEPSY scores in the intervention group. This progression can be explained by the distribution of students' profiles across the autism spectrum, as suggested by the correlation between SRS raw score and the two NEPSY subtests in this group. Although the difference was not statistically significant, the group using *KidLearn* displayed a lower raw SRS score relative to the control group. The intervention group included 6/14 students exceeding the clinical threshold for ASD, while the control group included 6/10 students in this case. This assumption is further supported by the fact that the NEPSY scores in the control group have decreased between pre- and post-intervention, even if the difference was not statistically significant (p = .075). In other words, the greater post-intervention NEPSY score observed in the intervention group might be related to the greater proportion of students with a lesser degree of autistic traits. These students might then be more likely to benefit from a test-retest learning effect on NEPSY subtests. As it is well-known, higher ASD severity is associated with more sociocognitive difficulties, including the processing of face-related stimuli (for review, Webb, et al., 2017). This interpretation is further supported by the psychometric qualities of the NEPSY subtests, for which test-retest reliability is adequate but perfectible at 21-days intervals (Brooks, et al., 2009; Korkman, et al., 2007). As our evaluations were only about a month apart, the results on the NEPSY subtests might have been influenced by test-retest learning effects.

Finally, we observed a pre-post increase of the KABC-II simultaneous processing score both in the intervention and the control group, suggesting that this effect was not related to the use of the *KidLearn* application. The use of tablets could account for this effect, but appeared unlikely in the face of other possibilities. Again, the psychometric qualities of the subtests provide an explanative way, given the short time between the pre- and the post-assessment (Kaufman, 2005; Kaufman & Kaufman, 2004). The simultaneous processing score was then likely to be confounded by learning effect with evaluations spaced one month apart. This might explain why we have observed an increase on the simultaneous processing score in both groups, and no significant progression for the sequential processing score, as the latter had a greater test-retest reliability coefficient.

Dealing with the diversity of special needs

The feasibility demonstration is equally important to the efficacy. We showed that ITS-based instruction of calculation skills is accessible to specialized classrooms for children with autism, with most children taking charge of their learning. During the intervention, a support need was expressed by some students with the highest calculation difficulties. The experimenter, the teacher and/or the school assistants aided those with greater difficulties in performing additions and subtractions. Depending on the profiles, the students might need occasional or permanent support from an adult. As it stood, a part of our sample had difficulties in using autonomously the *KidLearn* application, and needed the assistance of an adult to explain calculations and solution search processes. This situation was reminiscent of other studies that employed technologies with individuals with ASD as they often involved an instructor who could manage the administration of the intervention, or even initiate some of the individual's responses (*e.g.*, priming, reinforcement). In this regard, Chebli, *et al.* (2019) have recently evaluated a tablet-based application in which they minimized human interventions. They showed that this application was less effective than an educator-administered intervention. This study emphasized that technologies should not replace educators, but rather broaden the range of instructional tools available to them.

Another advantage of expert supervision is the feedback that experts provide about the user experience of the application. For instance, when the student uses up all attempts for a given trial, the only feedback they receive from the application alone is the correct set of banknotes and coins needed to obtain the expected price. Further explicit instruction that explains, for example, the calculation process and/or the causes of errors, might be beneficial for the student. Several studies have shown that some individual with ASD have difficulty in understanding and applying implicit rules (Geurts, et al., 2009; Klinger, et al., 2007): it was therefore recommended to make the instructions given to the students as explicit as possible (Bouck, et al., 2020; Cihak & Grim, 2008; Fletcher, et al., 2010; Hart Barnett & Cleary, 2015; Hsu, et al., 2014; Oswald, et al., 2016; Spooner, et al., 2019). If the students understand the cause of the error and are reminded of the calculation strategies, they will be likely to correct it in future attempts. Consequently, some adjustments in the interface and functionalities of the *KidLearn* application could be made for improving its design for students with ASD. However, the children enrolled in this study generally enjoyed to play with *KidLearn* and no major accessibility or ergonomic issue were reported.

Limitations and future studies

Despite promising results, this study has several limitations, which limits the scope of our conclusions while opening to future works. First, some limitations are mainly related to the fact that this pilot study was intended to explore the feasibility of ITS-based instruction with children with ASD: a small sample size and a small intervention duration. A larger sample size will increase the statistical power and provide stronger evidence related to the intervention effects. Also, the intervention duration was relatively small, and we did not examine the maintenance of effects after the intervention. Encouraging results were yet obtained within a short period of time, but the question of whether these effects were maintained after the intervention should be investigated.

Another limitation is related to the choice of the control condition. Including a control group is required for successfully assessing the outcomes of an intervention, but its design must be adapted to the research context and the study's objectives. In behavioral intervention trials, many factors may contribute to the generation of placebo effects, such as resources provided or attention given to the participants. Then, *"the control group must be designed to* *create the nonspecific conditions, such as expectancy, social support, and attention, considered necessary to generate placebo effects"* (Lindquist, et al., 2007, p. 215). The active control condition allows for controlling contextual effects related to the occurrence of an intervention, but cannot explore the added value of our ITS approach. The control group design could be improved to explore the specific contribution of ITS in comparison with other instructional practices, with, for example, a second control condition involving either *KidLearn* without ZPDES (with a linear predefined teaching sequence within the ITS) and/or traditional math instruction by a human teacher (usual instruction).

Future studies should move forward to reliably assess both general effects related to digital educational interventions for children with ASD, and specific effects related to the ITS feature in comparison with other instructional practices. A three-armed trial with one experimental condition and two control conditions (algorithm providing a linear predefined path, and usual instruction) would allow for assessing the specific effects of ITS-based *KidLearn* intervention with students with ASD and/or ID in specialized classrooms. Standardized measurements of mathematical skills would allow for assessing global mathematical skills. Also, the skills' generalization and maintenance would be explored through respectively, a more ecologic transfer task and an additional follow-up assessment. This design requires a large-scale recruitment in order to have sufficient statistical power for providing high-quality evidence. Such large-scale recruitment will also allow for analyzing the results depending on the diagnosis and then, highlight ASD-specific outcomes related to the use of ITS-based instruction.

Conclusion

The present pilot study supported the relevance of the *KidLearn* application equipped with the ZPDES algorithm to instruct basic mathematical abilities to secondary students with ASD and/or ID. Significant learning progress was observed in less than one month of using the *KidLearn* application. These results call for further studies over a longer intervention period for examining whether the students continue to progress, and even master basic calculations. Other variants of algorithms could be tested to determine the added value of the ZPDES algorithm in fostering learning gains and motivation in students. We should keep in mind that this pilot study has been conducted on a reduced number of students, thus limiting the generalization of the results. Further investigations including a larger sample of students is required to confirm our preliminary results.

Finally, we showed that ITS can be successfully used with secondary students with ASD and/or ID and be a promising way to create a new generation of psycho-educational interventions for individuals with ASD. The ZPDES algorithm proposed to select activities that adapt to the level and the progress of each learner, and then allowed a high degree of curriculum personalization. This characteristic is also an essential ingredient in interventions targeting individuals with ASD and/or ID (*e.g.*, Schreibman, et al., 2015; Wong, et al., 2014). These preliminary results support a new research avenue in which students with ASD and/or ID in specialized classrooms could benefit from a personalized technology-based intervention in the same way as typically-developing students enrolled in regular classrooms.

Reference List

Aljameel, S., O'Shea, J., Crockett, K., Latham, A., & Kaleem, M. (2017). Development of An Arabic Conversational Intelligent Tutoring System for Education of Children with ASD. *2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, 24–29. https://doi.org/10.1109/CIVEMSA.2017.7995296

Aljameel, S., O'Shea, J., Crockett, K., Latham, A., & Kaleem, M. (2019). LANA-I: An Arabic Conversational Intelligent Tutoring System for Children with ASD. In K. Arai, R. Bhatia, & S. Kapoor (Eds.), *CompCom 2019: Intelligent Computing* (pp. 498–516). Sringer. https://doi.org/10.1007/978-3-030-22871-2_34

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition*. American Psychiatric Publishing, Inc. https://doi.org/10.1176/appi.books.9780890425596.893619

Barendse, E. M., Hendriks, M. P., Jansen, J. F., Backes, W. H., Hofman, P. A., Thoonen, G., Kessels, R. P., & Aldenkamp, A. P. (2013). Working memory deficits in high-functioning adolescents with autism spectrum disorders: neuropsychological and neuroimaging correlates. *Journal of Neurodevelopmental Disorders*, *5*(1), 1–11. https://doi.org/10.1186/1866-1955-5-14

Bernaud, J.-L. (2014). *Méthodes de tests et questionnaires en psychologie*. Dunod.

Bouck, E. C., Mathews, L. A., & Peltier, C. (2020). Virtual Manipulatives: A Tool to Support Access and Achievement With Middle School Students With Disabilities. *Journal of Special Education Technology*, *35*(1), 51–59. https://doi.org/10.1177/0162643419882422

Brennan, L. M., Shaw, D. S., Dishion, T. J., & Wilson, M. (2012). Longitudinal predictors of school-age academic achievement: Unique contributions of toddler-age aggression, oppositionality, inattention, and hyperactivity. *Journal of Abnormal Child Psychology*, *40*(8), 1289–1300. https://doi.org/10.1007/s10802-012-9639-2

Brooks, B. L., Sherman, E. M. S., & Strauss, E. (2009). NEPSY-II: A Developmental Neuropsychological Assessment, Second Edition. *Child Neuropsychology*, *16*(1), 80–101. https://doi.org/10.1080/09297040903146966

Brosnan, M., Johnson, H., Grawemeyer, B., Chapman, E., Antoniadou, K., & Hollinworth, M. (2016). Deficits in metacognitive monitoring in mathematics assessments in learners with autism spectrum disorder. *Autism*, *20*(4), 463–472. https://doi.org/10.1177/1362361315589477

Bullen, J. C., Swain Lerro, L., Zajic, M., McIntyre, N., & Mundy, P. (2020). A Developmental Study of Mathematics in Children with Autism Spectrum Disorder, Symptoms of Attention Deficit Hyperactivity Disorder, or Typical Development. *Journal of Autism and Developmental Disorders*, *50*(12), 4463–4476. https://doi.org/10.1007/s10803-020-04500-9

Chebli, S. S., Lanovaz, M. J., & Dufour, M.-M. (2019). Comparison of Tablet-Delivered and Instructor-Delivered Teaching on Receptive Identification in Children With Autism Spectrum Disorders. *Journal of Special Education Technology*, *34*(1), 55–67. https://doi.org/10.1177/0162643418781300

Chiang, H. M., & Lin, Y. H. (2007). Mathematical ability of students with Asperger syndrome and high-functioning autism: A review of literature. *Autism*, *11*(6), 547–556. https://doi.org/10.1177/1362361307083259

Cihak, D. F., & Grim, J. (2008). Teaching students with autism spectrum disorder and moderate intellectual disabilities to use counting-on strategies to enhance independent purchasing skills. *Research in Autism Spectrum Disorders*, *2*(4), 716–727. https://doi.org/10.1016/j.rasd.2008.02.006

Clément, B. (2018). *Adaptive Personalization of Pedagogical Sequences using Machine Learning*. (doctoral dissertation) University of Bordeaux.

Clement, B., Roy, D., Oudeyer, P.-Y., & Lopes, M. (2013). Multi-Armed Bandits for Intelligent Tutoring Systems. *Journal of Educational Mining*, *7*(2), 20–48. http://arxiv.org/abs/1310.3174

Constantino, J. N., & Gruber, C. P. (2012). *Social Responsiveness Scale - Second version (SRS-II)*. Western Psychological Services.

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, *3*(2), 63–68. https://doi.org/10.1016/j.tine.2013.12.001

Crone, D. L., Bode, S., Murawski, C., & Laham, S. M. (2018). The Socio-Moral Image Database (SMID): A novel stimulus set for the study of social, moral and affective processes. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0190954

Danielsson, H., Henry, L., Rönnberg, J., & Nilsson, L.-G. (2010). Executive functions in individuals with intellectual disability. *Research in Developmental Disabilities*, *31*(6), 1299–1304. https://doi.org/10.1016/j.ridd.2010.07.012

Dattolo, A., & L. Luccio, F. (2017). Accessible and UsableWebsites and Mobile Applications for People with Autism Spectrum Disorders: a Comparative Study. *ICST Transactions on Ambient Systems*, *4*(13), 152549. https://doi.org/10.4108/eai.17-5-2017.152549

Dattolo, A., & Luccio, F. L. (2017). A Review of Websites and Mobile Applications for People with Autism Spectrum Disorders: Towards Shared Guidelines. In O. Gaggi, P. Manzoni, C. Palazzi, A. Bujari, & J. M. Marquez-Barja (Eds.), *GOODTECHS 2016: Smart Objects and Technologies for Social Good* (pp. 264–273). Springer. https://doi.org/10.1007/978-3-319-61949-1_28

Delmas, A., Clement, B., Oudeyer, P. Y., & Sauzéon, H. (2018). Fostering Health Education With a Serious Game in Children With Asthma: Pilot Studies for Assessing Learning Efficacy and Automatized Learning Personalization. *Frontiers in Education*, *3*. https://doi.org/10.3389/feduc.2018.00099

Fletcher, D., Boon, R. T., & Cihak, D. F. (2010). Effects of the touchmath program compared to a number line strategy to teach addition facts to middle school students with moderate intellectual disabilities. *Education and Training in Autism and Developmental Disabilities*, *45*(3), 449–458.

Fleury, V. P., Hedges, S., Hume, K., Browder, D. M., Thompson, J. L., Fallin, K., El Zein, F., Reutebuch, C. K., & Vaughn, S. (2014). Addressing the Academic Needs of Adolescents With Autism Spectrum Disorder in Secondary Education. *Remedial and Special Education*, *35*(2), 68–79. https://doi.org/10.1177/0741932513518823

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, *47*(6), 1539–1552. https://doi.org/10.1037/a0025510

Geurts, H. M., Corbett, B., & Solomon, M. (2009). The paradox of cognitive flexibility in autism. *Trends in Cognitive Sciences*, *13*(2), 74–82. https://doi.org/10.1016/j.tics.2008.11.006

Gevarter, C., Bryant, D. P., Bryant, B., Watkins, L., Zamora, C., & Sammarco, N. (2016). Mathematics Interventions for Individuals with Autism Spectrum Disorder: A Systematic Review. *Review Journal of Autism and Developmental Disorders*, *3*(3), 224–238. https://doi.org/10.1007/s40489-016-0078-9

Gioia, G. A., Isquith, P. K., Retzlaff, P. D., & Espy, K. A. (2002). Confirmatory factor analysis of the Behavior Rating Inventory of Executive Function (BRIEF) in a clinical sample. *Child Neuropsychology*, *8*(4), 249–257. https://doi.org/10.1076/chin.8.4.249.13513

Goodall, C. (2015). How do we create ASD-friendly schools? A dilemma of placement. *Support for Learning*, *30*(4), 305–326. https://doi.org/10.1111/1467-9604.12104

Grégoire, J. (2009). *L'examen clinique de l'intelligence de l'enfant: fondements et pratique du WISC-IV* (2nd ed.). Editions Mardaga.

Grynszpan, O., Martin, J.-C., & Nadel, J. (2008). Multimedia interfaces for users with high functioning autism: An empirical investigation. *International Journal of Human-Computer Studies*, *66*(8), 628–639. https://doi.org/10.1016/j.ijhcs.2008.04.001

Hammer, D., Melhuish, E., & Howard, S. J. (2017). Do aspects of social, emotional and behavioural development in the pre-school period predict later cognitive and academic attainment? *Australian Journal of Education*, *61*(3), 270–287. https://doi.org/10.1177/0004944117729514

Hart Barnett, J. E., & Cleary, S. (2015). Review of evidence-based mathematics interventions for students with autism spectrum disorders. *Education and Training in Autism and Developmental Disabilities*, *50*(2), 172–185.

Hill, E. L. (2004). Executive dysfunction in autism. *Trends in Cognitive Sciences*, *8*(1), 26–32. https://doi.org/10.1016/j.tics.2003.11.003

Hsu, G.-L., Tang, J.-C., & Hwang, W.-Y. (2014). Effects of extending the one-more-than technique with the support of a mobile purchasing assistance system. *Research in Developmental Disabilities*, *35*(8), 1809–1827. https://doi.org/10.1016/j.ridd.2014.04.004

Jahromi, L. B., Bryce, C. I., & Swanson, J. (2013). The importance of self-regulation for the school and peer engagement of children with high-functioning autism. *Research in Autism Spectrum Disorders*, *7*(2), 235–246. https://doi.org/10.1016/j.rasd.2012.08.012

Jones, C. R. G., Happé, F., Golden, H., Marsden, A. J. S., Tregay, J., Simonoff, E., Pickles, A., Baird, G., & Charman, T. (2009). Reading and arithmetic in adolescents with autism spectrum disorders: Peaks and dips in attainment. *Neuropsychology*, *23*(6), 718–728. https://doi.org/10.1037/a0016360

Kaufman, A. S., & Kaufman, N. L. (2004). *Manual for the Kaufman Assessment Battery for Children: Second Edition (KABC-II).* AGS Publishing.

Kaufman, A. S., Lichtenberger, E. O., Pletcher-Janzen, E., & Kaufman, N. L. (2005). *Essentials of KABC-II assessment.* Wiley.

Keen, D. (2009). Engagement of Children With Autism in Learning. *Australasian Journal of Special Education*, *33*(2), 130–140. https://doi.org/10.1375/ajse.33.2.130

Keen, D., Webster, A., & Ridley, G. (2016). How well are children with autism spectrum disorder doing academically at school? An overview of the literature. *Autism*, *20*(3), 276–294. https://doi.org/10.1177/1362361315580962

Kenworthy, L., Anthony, L. G., Naiman, D. Q., Cannon, L., Wills, M. C., Luong-Tran, C., Werner, M. A., Alexander, K. C., Strang, J., Bal, E., Sokoloff, J. L., & Wallace, G. L. (2014). Randomized controlled effectiveness trial of executive function intervention for children on the autism spectrum. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *55*(4), 374–383. https://doi.org/10.1111/jcpp.12161

Kiep, M., & Spek, A. A. (2017). Executive functioning in men and women with an autism spectrum disorder. *Autism Research*, *10*(5), 940–948. https://doi.org/10.1002/aur.1721

Kiru, E. W., Doabler, C. T., Sorrells, A. M., & Cooc, N. A. (2018). A Synthesis of Technology-Mediated Mathematics Interventions for Students With or at Risk for Mathematics Learning Disabilities. *Journal of Special Education Technology*, *33*(2), 111–123. https://doi.org/10.1177/0162643417745835

Klinger, L. G., Klinger, M. R., & Pohlig, R. L. (2007). Implicit learning impairments in autism spectrum disorders. In C. N. Vizcaino, J. M. Perez, M. L. Comi, & P. M. Gonzalez (Eds.), *New Developments in Autism: The Future in Today* (pp. 76–102). Jessica Kingsley Publishers.

Knight, V., McKissick, B. R., & Saunders, A. (2013). A Review of Technology-Based Interventions to Teach Academic Skills to Students with Autism Spectrum Disorder.

Journal of Autism and Developmental Disorders, *43*(11), 2628–2648. https://doi.org/10.1007/s10803-013-1814-y

Koegel, L. K., Singh, A. K., & Koegel, R. L. (2010). Improving Motivation for Academics in Children with Autism. *Journal of Autism and Developmental Disorders*, *40*(9), 1057–1066. https://doi.org/10.1007/s10803-010-0962-6

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-Second Edition (NEPSY-II)*. Harcourt Assessment.

Kover, S. T., & Atwood, A. K. (2013). Establishing Equivalence: Methodological Progress in Group-Matching Design and Analysis. *American Journal of Intellectual Developmental Disabilities*, *118*(1), 3–15. https://doi.org/10.1352/1944-7558-118.1.3

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review. *Review of Educational Research*, *86*(1), 42–78. https://doi.org/10.3102/0034654315581420

Kurth, J., & Mastergeorge, A. M. (2010). Individual Education Plan Goals and Services for Adolescents With Autism: Impact of Age and Educational Setting. *The Journal of Special Education*, *44*(3), 146–160. https://doi.org/10.1177/0022466908329825

Lindquist, R., Wyman, J. F., Talley, K. M. C., Findorff, M. J., & Gross, C. R. (2007). Design of control-group conditions in clinical trials of behavioral interventions. *Journal of Nursing Scholarship*, *39*(3), 214–221. https://doi.org/10.1111/j.1547-5069.2007.00171.x

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, *106*(4), 901–918. https://doi.org/10.1037/a0037123

Maras, K., Gamble, T., & Brosnan, M. (2019). Supporting metacognitive monitoring in mathematics learning for young people with autism spectrum disorder: A classroom-based study. *Autism*, *23*(1), 60–70. https://doi.org/10.1177/1362361317722028

Mazon, C., Fage, C., & Sauzéon, H. (2019). Effectiveness and usability of technology-based interventions for children and adolescents with ASD: A systematic review of reliability, consistency, generalization and durability related to the effects of intervention. *Computers in Human Behavior*, *93*, 235–251. https://doi.org/10.1016/j.chb.2018.12.001

Meltzer, L. (2010). *Promoting Executive Function in the Classroom (What Works for Special Learner)*. Guilford Press.

Meltzer, L., & Krishnan, K. (2007). Executive function difficulties and learning disabilities: Understandings and Misunderstandings. In L. Meltzer (Ed.), *Executive function in education: From theory to practice* (pp. 77–105). Guilford Press.

Metcalfe, J., & Kornell, N. (2005). A Region of Proximal Learning model of study time allocation. *Journal of Memory and Language*, *52*(4), 463–477. https://doi.org/10.1016/j.jml.2004.12.001

Milani, S., Fan, Z., Gulati, S., Nguyen, T., Fang, F., & Yadav, A. (2020). Intelligent Tutoring Strategies for Students with Autism Spectrum Disorder : A Reinforcement Learning Approach. *The 2020 CMU Symposium on Artificial Intelligence and Social Good*.

Mondragon, A. L., Nkambou, R., & Poirier, P. (2016). Evaluating the Effectiveness of an Affective Tutoring Agent in Specialized Education. In K. Verbert, M. Sharples, & T. Klobucar (Eds.), *EC-TEL 2016: Adaptive and Adaptable Learning* (pp. 446–452). Springer.

Morgan, P. L., Farkas, G., Hillemeier, M. M., Hammer, C. S., & Maczuga, S. (2015). 24-Month-Old Children With Larger Oral Vocabularies Display Greater Academic and Behavioral Functioning at Kindergarten Entry. *Child Development*, *86*(5), 1351–1370. https://doi.org/10.1111/cdev.12398

Morsanyi, K., van Bers, B. M. C. W., McCormack, T., & McGourty, J. (2018). The prevalence of specific learning disorder in mathematics and comorbidity with other developmental disorders in primary school-age children. *British Journal of Psychology*, *109*(4), 917–940. https://doi.org/10.1111/bjop.12322

Naglieri, J. A., & Chambers, K. M. (2009). Psychometric issues and current scales for assessing autism spectrum disorders. In J. A. Naglieri & S. Ozonoff (Eds.), *Assessment of autism spectrum disorders.* (pp. 55–90). Guilford Press.

Nkambou, R., Bourdeau, J., & Mizoguchi, R. (2010). Introduction: What are Intelligent Tutoring Systems, and why this book? In R. Nkambou, R. Mizoguchi, & J. Bourdeau (Eds.), *Advances in Intelligent Tutoring Systems* (Vol. 308, pp. 1–11). Springer Science & Business Media.

Oswald, T. M., Beck, J. S., Iosif, A.-M., McCauley, J. B., Gilhooly, L. J., Matter, J. C., & Solomon, M. (2016). Clinical and Cognitive Characteristics Associated with Mathematics Problem Solving in Adolescents with Autism Spectrum Disorder. *Autism Research*, *9*(4), 480–490. https://doi.org/10.1002/aur.1524

Oudeyer, P.-Y., Gottlieb, J., & Lopes, M. (2016). Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. In B. Studer & S. Knecht (Eds.), *Progress in Brain Research* (Vol. 229, pp. 257–284). Elsevier.

Ozonoff, S., & Schetter, P. L. (2007). Executive dysfunction in autism spectrum disorders: from research to practice. In L. Meltzer (Ed.), *Executive function in education: From theory to practice* (pp. 133–160). Guilford Press.

Panerai, S., Tasca, D., Ferri, R., Genitori D'Arrigo, V., & Elia, M. (2014). Executive Functions and Adaptive Behaviour in Autism Spectrum Disorders with and without Intellectual Disability. *Psychiatry Journal*, *2014*, 941809. https://doi.org/10.1155/2014/941809

Pavlov, N. (2014). User Interface for People with Autism Spectrum Disorders. *Journal of Software Engineering and Applications*, *7*(2), 128–134. https://doi.org/10.4236/jsea.2014.72014

Pett, M. A. (2015). *Nonparametric testing for Health Care Research: Statistics for Small sample sizes and Unusual Distributions* (Sage Publications (ed.)).

Pitchford, N. J., Kamchedzera, E., Hubber, P. J., & Chigeda, A. L. (2018). Interactive Apps Promote Learning of Basic Mathematics in Children With Special Educational Needs and Disabilities. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00262

Roelofs, R. L., Visser, E. M., Berger, H. J. C., Prins, J. B., Van Schrojenstein Lantman-De Valk, H. M. J., & Teunisse, J. P. (2015). Executive functioning in individuals with intellectual disabilities and autism spectrum disorders. *Journal of Intellectual Disability Research*, *59*(2), 125–137. https://doi.org/10.1111/jir.12085

Rose, E., & Meyer, A. (2006). *A practical reader in unvisersal design for learning*. Harvard Education Press.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Samuels, W. E., Tournaki, N., Blackman, S., & Zilinski, C. (2016). Executive functioning predicts academic achievement in middle school: A four-year longitudinal study. *Journal of Educational Research*, *109*(5), 478–490. https://doi.org/10.1080/00220671.2014.979913

Sarma, B. H. S., & Ravindran, B. (2007). Intelligent Tutoring Systems using Reinforcement Learning to teach Autistic Students. In A. Venkatesh, T. Gonsalves, A. Monk, & K. Buckner (Eds.), *Home Informatics and Telematics: ICT for The Next Billion* (Vol. 241, pp. 65–78). Springer. https://doi.org/10.1007/978-0-387-73697-6_5

Schanding, G. T., Nowell, K. P., & Goin-Kochel, R. P. (2012). Utility of the Social Communication Questionnaire-Current and Social Responsiveness Scale as Teacher-Report Screening Tools for Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *42*(8), 1705–1716. https://doi.org/10.1007/s10803-011-1412-9

Schreibman, L., Dawson, G., Stahmer, A. C., Landa, R., Rogers, S. J., McGee, G. G., Kasari, C., Ingersoll, B., Kaiser, A. P., Bruinsma, Y., McNerney, E., Wetherby, A., & Halladay, A. (2015). Naturalistic Developmental Behavioral Interventions: Empirically Validated Treatments for Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *45*(8), 2411–2428. https://doi.org/10.1007/s10803-015-2407-8

Soorya, L. V., Siper, P. M., Beck, T., Soffes, S., Halpern, D., Gorenstein, M., Kolevzon, A., Buxbaum, J., & Wang, A. T. (2015). Randomized Comparative Trial of a Social Cognitive Skills Group for Children With Autism Spectrum Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*(3), 208-216.e1. https://doi.org/10.1016/j.jaac.2014.12.005

Spaniol, M. M., Shalev, L., Kossyvaki, L., & Mevorach, C. (2018). Attention Training in Autism as a Potential Approach to Improving Academic Performance: A School-

Based Pilot Study. *Journal of Autism and Developmental Disorders*, *48*(2), 592–610. https://doi.org/10.1007/s10803-017-3371-2

Spooner, F., Root, J. R., Saunders, A. F., & Browder, D. M. (2019). An Updated Evidence-Based Practice Review on Teaching Mathematics to Students With Moderate and Severe Developmental Disabilities. *Remedial and Special Education*, *40*(3), 150–165. https://doi.org/10.1177/0741932517751055

Stordeur, C., Boele, A., Peyre, H., Delorme, R., & Acquaviva, E. (2019). Psychometric properties of the French Version of the Social Responsiveness Scale in autism spectrum disorder with or without attention deficit hyperactivity disorder. *L'Encéphale*, *45*(4), 285–289. https://doi.org/10.1016/j.encep.2018.08.004

Vallerand, R. J., Blais, M. R., Brière, N. M., & Pelletier, L. G. (1989). Construction et validation de l'échelle de motivation en éducation (EME). *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement*, *21*(3), 323–349. https://doi.org/10.1037/h0079855

vanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Vickers, A. J. (2005). Parametric versus non-parametric statistics in the analysis of randomized trials with non-normally distributed data. *BMC Medical Research Methodology*, *5*(1), 35. https://doi.org/10.1186/1471-2288-5-35

Vygotsky, L. (1980). *Mind in society: the development of higher psychological processes*. Harvard University Press.

Webb, S. J., Neuhaus, E., & Faja, S. (2017). Face Perception and Learning in Autism Spectrum Disorders. *Quarterly Journal of Experimental Psychology*, *70*(5), 970–986. https://doi.org/10.1080/17470218.2016.1151059

Wechsler, D. (2003). *Weschler Intelligence Scale for Children-Fourth Version (WISC-IV).*

Wei, X., Christiano, E. R., Yu, J. W., Wagner, M., & Spiker, D. (2015). Reading and math achievement profiles and longitudinal growth trajectories of children with an autism spectrum disorder. *Autism*, *19*(2), 200–210. https://doi.org/10.1177/1362361313516549

Wong, C., Odom, S. L., Hume, K. A., Cox, A. W., Fettig, A., Kucharczyk, S., Brock, M. E., Plavnick, J. B., Fleury, V. P., & Schultz, T. R. (2015). Evidence-Based Practices for Children, Youth, and Young Adults with Autism Spectrum Disorder: A Comprehensive Review. *Journal of Autism and Developmental Disorders*, *45*(7), 1951–1966. https://doi.org/10.1007/s10803-014-2351-z

Zaidman-Zait, A., Mirenda, P., Szatmari, P., Duku, E., Smith, I. M., Zwaigenbaum, L., Vaillancourt, T., Kerns, C., Volden, J., Waddell, C., Bennett, T., Georgiades, S., Ungar, W. J., & Elsabbagh, M. (2021). Profiles and Predictors of Academic and Social School Functioning among Children with Autism Spectrum Disorder. *Journal of*

Clinical Child and Adolescent Psychology, *50*(5), 656–668. https://doi.org/10.1080/15374416.2020.1750021

Appendix

Profile questionnaire (translated from French)

1. **Are you :** a) a girl b) a boy
2. **Have you ever used money ?**
   a) I have used money before and I like to do it.
   b) I have used money before, but I don’t like to do it
   c) I never used money before and I'd like to learn how to do it.
   d) I never used money before, but I don’t want to learn how to do it.
3. **Do you like to do calculations ?** a) No b) Yes
4. **What devices did you use at home ?**
   a) A computer
   b) A tablet
   c) A smartphone
   d) A handheld game console
   e) A home game console
5. **How often did you use these screens ?**
   a) Never
   b) Only on week-ends
   c) A few times during the week
   d) Every day
6. **At home, I choose the activities I do :**
   a) Not at all
   b) Not often
   c) Sometimes yes, sometimes no
   d) Often

e) All the time

7. In general, I choose the food I eat :

a) Not at all

b) Not often

c) Sometimes yes, sometimes no

d) Often

e) All the time

8. I choose who my friends are :

a) Not at all

b) Not often

c) Sometimes yes, sometimes no

d) Often

e) All the time

9. I choose the clothes I wear :

a) Not at all

b) Not often

c) Sometimes yes, sometimes no

d) Often

e) All the time

Author Note

Cécile Mazon: https://orcid.org/0000-0003-3959-2712

Benjamin Clément: https://orcid.org/0000-0002-7966-3661

Didier Roy: https://orcid.org/0000-0002-0904-0506

Pierre-Yves Oudeyer: https://orcid.org/0000-0002-1277-130X

Hélène Sauzéon: https://orcid.org/0000-0001-5781-9891

Acknowledgements: The authors express their gratitude to the children, their families and school staffs who participated to the study. The authors gratefully acknowledge the academy and the rectorate of Bordeaux as well as the head of the special education sector, for their support in the project and the recruitment of special education classrooms. The authors also acknowledge the financial support provided by the Orange Foundation and Inria Bordeaux Sud-Ouest. Finally, the authors thank Alexander Ten for his help in the final revision of the article.

Conflicts of Interests: The authors declare that they have no conflict of interest.

Authors contributions: CM: Conceptualization; Methodology; Investigation; Formal analysis; Writing – original draft; BC: Conceptualization; Software; Resources; Data curation; DR: Conceptualization; Software; Resources; Methodology; PYO: Conceptualization; Validation; Supervision; Project administration; Funding acquisition; HS: Conceptualization; Methodology; Validation; Writing – Review & Editing; Supervision; Project administration; Funding acquisition

Correspondence concerning this article should be addressed to Cécile Mazon or Hélène Sauzéon, Flowers project-team, Inria center of University of Bordeaux, 200 avenue de la Vieille Tour, 33405 Talence Cedex, France, helene.sauzeon@inria.fr and/or cecile.mazon@inria.fr