UNIVERSITY OF SZEGED
DOCTORAL SCHOOL OF EDUCATION

DE VAN VO

ASSESSING INDUCTIVE REASONING, SCIENTIFIC REASONING AND SCIENCE MOTIVATION: CROSS-SECTIONAL STUDIES IN VIETNAMESE CONTEXT

DOCTORAL DISSERTATION

SUPERVISOR:
Prof. Dr. BENŐ CSAPÓ

SZEGED, HUNGARY, 2022

THE CONTEXT AND STRUCTURE OF THE DISSERTATION

In Vietnam, the national educational program aims to develop students' thinking proficiency through implicit discipline curricula. However, no empirical studies were broadly conducted to explore general reasoning skills in Vietnamese context. With this as a backdrop, the main goals of our investigations are to validate the adapted test instruments for assessing reasoning skills and science motivation and to explore the patterns of students' reasoning and science motivation across grade levels. The cross-sectional investigations can offer useful information about how well learning science and teaching thinking skills can be integrated into subject-specific areas at different grade levels. This may be meaningful in practice vis-à-vis boosting students' reasoning capacities and motivation and in proposing improved programmes in future. Our study includes a review of the findings from previous empirical studies and a set of three cross-sectional studies to assess reasoning abilities and science motivation of secondary school students in Vietnam.

The dissertation is organized into eight chapters. Chapter 1 highlights general core facets of research problems in the project. It provides main points from former studies in reasoning and science motivation in educational settings around the world as well as introduces the study context in Vietnam in line with the aims of our research project. Chapter 2 presents the main characteristics and trends in measuring IR and SR around the world through a systematic review with 63 empirical studies from 1997 to 2020. In addition, motivational factors and assessment of science motivation were underlined in this chapter. In Chapter 3, we discuss the main methods applied in three empirical studies. The test instruments are discussed in this chapter, including the IR test, SR test, CVS test in physics, science motivation questionnaire, student motivation and attitude toward physics learning, and the background questionnaire. Moreover, we summarized the main methods for data analysis and interpretation such as the principal component analysis (PCA), confirmatory factor analysis (CFA), Rasch model measurement, differential item difficulty (DIF), path analysis and multiple linear regression models with Bayesian model averaging (BMA) approach.

Chapter 4 illustrates the first empirical study that aims to investigate the development of IR and patterns of students' motivation across the 5^{th}, 7^{th}, 9^{th}, and 11^{th} grades. The initial findings showed that students' performance improved steadily grade by grade, whereas their motivation toward learning science gradually fell through the grade levels. The main factors contributing to individual IR capacity and science motivation are also discussed in this chapter. The findings of the second study are reported in Chapter 5, which showed the developmental curves of IR, SR capacities and SM in the 6^{th}, 8^{th}, 10^{th}, and 11^{th} graders. The developmental curves in IR and SR were simulated across grade cohorts, while the students' SM seems to reduce in the upper grade cohorts. Furthermore, the predictive role of IR, SR and SM on children's STEM achievement are discussed in this chapter. Chapter 6 details the third empirical study which considers assessing scientific reasoning in CVS in physics in secondary school students. This chapter presents the development and validity of the CVS test to measure three subskills of CVS related to content knowledge in basic physics (mechanics, thermodynamics, and electricity). In general, students' CVS proficiency progressed during secondary education years, but their motivation toward physics learning slightly reduced during secondary education level. Relationships between CVS, motivation and content knowledge in physics were also examined in this study.

Chapter 7 discusses the effects of media delivery by comparing the online and PP administration modes in students' performance on the cognitive tests across these empirical studies. The examinations of the equivalence of dual test versions were conducted in both internal structural factors and DIF analyses. We employed classical statistics and Rasch model measurement to compare psychometric properties of the tests and students' score results at the test, task, and item levels. Finally, Chapter 8 summarizes the general conclusion and discussion across these investigations. The limitations and future directions are also discussed in the last chapter.

THEORETICAL BACKGROUND

Assessment plays an essential role in the teaching and learning process as well as in school administration. Assessment involves the process of measuring, collecting, and using evidence about the outcomes of students' learning, provides feedback for students in their learning processes, and helps teachers and other stakeholders make evidence-based decisions (Hattie & Timperley, 2007). The technological transformation makes it possible to assess wider-ranging student performance, including assessing core reasoning skills (Osborne, 2013).

Reasoning can broadly be defined as the goal-driven process of drawing conclusions which inform problem-solving and decision-making efforts (Leighton & Sternberg, 2004). Amongst reasoning forms, IR and SR are mostly investigated in school settings due to its important roles in learning school subjects and the feasibility to enhance these skills in educational contexts. IR is one of the seven primary mental abilities that contributes to intelligent behaviour (Kinshuk et al., 2006). It refers to a cognitive process in which particular facts or individual cases are gathered to establish a general conclusion (Adey & Csapó, 2012; Sternberg & Sternberg, 2012). Researchers have found that IR is closely tied to problem-solving skills (Csapó, 1997; Molnár et al., 2013; Schweizer et al., 2013) and plays an important role in learning most school subjects, such as science (Adey, & Csapó, 2012; Hamers et al., 1998), mathematics (Nunes & Csapó, 2011) and foreign languages (Nikolov & Csapó, 2018; Soodmand Afshar et al., 2014) as well as other subjects. To assess IR ability, there are four groups of problem tasks: analogies, series completion, classifications and matrices (Adey & Csapó, 2012; Sternberg & Sternberg, 2012). The analogies tasks, constructed of the materials in geometry and figure, are the most intensively applied in a variety of IR tests.

Empirical studies have confirmed that children's IR ability develops grade by grade, but growth rates seemed to be different across grade levels (Díaz-Morales & Escribano, 2013; Molnár et al., 2013; Muniz et al., 2012). Muniz et al. (2012) showed that students improved their IR capacity in primary schools (the 1^{st} to 5^{th} grades) with a steady development in the period of the 3^{rd}–11^{th} grades, but the most rapid development was noted between the 6^{th} and 7^{th} grades (12–14 years) (Díaz-Morales & Escribano, 2013; Molnár et al., 2013). As regards gender differences, these studies have shown inconsistent findings. Several studies (Blum et al., 2016; Jeotee, 2012; Kambeyo & Wu, 2018; Kyllonen et al., 2019; Tairab, 2015; Tekkaya & Yenilmez, 2006; Venville & Oliver, 2015) have demonstrated a significant difference in favour of males. In contrast, there were studies which found out that girls attained higher scores than boys did on the IR test (e.g. Díaz-Morales & Escribano, 2013). Meanwhile, other research has indicated no significant differences in reasoning skills regarding gender (Molnár, 2011; Salihu et al., 2018). These divergent results may be linked to an underlying factor of culture.

With a growing emphasis on science education, a further form of reasoning has become a focal research topic. The term scientific reasoning (SR) has been used as a domain-specific approach in science subjects. According to Lawson (2009), SR is considered as one of the foundational pillars of scientific literacy, along with content knowledge of science. SR is defined as an active procedure of interrelating a series of cognitive and metacognitive processes to generate, test, and revise theories and hypotheses (Zimmerman, 2007). SR plays a central role in science subjects in general and in separate subjects (mathematics, physics, biology and chemistry) (Zimmerman, 2000). To measure SR capacity, some kinds of tasks have been proposed, so-called conservation, control of variables, proportions and ratios, probability, correlational reasoning, hypothetical-deductive reasoning and combinatorial reasoning (Adey & Csapó, 2012; Han, 2013). Some IR tasks which the elements are constructed of science content have considered as a kind of scientific reasoning problem tasks.

Students' SR capacity increased gradually from one grade to the next through the general education (Korom et al., 2017; Tairab, 2015), but the growth rate tended to decrease after the 9th grade (14 years) (Ding, 2018; Kwon & Lawson, 2000). As regards gender differences in scientific reasoning, there have been inconsistent findings among previous studies. Some have found that boys performed better than girls on the SR tests (Tairab, 2015; Tekkaya & Yenilmez, 2006), but others have reported no significant differences in SR between males and females (Mayer et al., 2014; Piraksa et al., 2014; Thuneberg et al., 2015).

Motivation toward learning results from the relative dynamics of dispositional and contextual variables (Pintrich & Schunk, 2002). Academic motivation is related to students' goals, the intrinsic and extrinsic nature of motivation, students' beliefs about their competencies and students' perceived evaluation of academic tasks (Garcia & Pintrich, 1995). Numerous empirical studies (e.g. Chan & Norlizah, 2018; Tuan et al., 2005) have demonstrated that students with higher motivation were likely to perform better in science. The five-year panel investigation by Hwang et al. (2016) also found a longitudinal causal relationship between school achievement and self-efficacy.

Some instruments have been recommended in the literature for measuring SM. For example, Glynn and his colleagues (2009) developed the Science Motivation Questionnaire II to examine five motivational factors: intrinsic motivation, self-determination, self-efficacy, career motivation and grade motivation. Additionally, Tuan et al. (2005) proposed the students' motivation towards science learning (SMTSL) questionnaire, which combines constructivist learning and motivation and uses a five-point Likert format. It covers self-efficacy, active learning strategies, science learning value, performance goals, achievement goals and learning environment stimulation.

The findings regarding gender in motivation were divergent and depended on both different motivational components and particular contexts. For instance, boys did not differ from girls on the intrinsic and extrinsic motivation, self-determination or self-efficacy scales (Britner, 2008; Glynn et al., 2009; Zeyer, 2010; Zeyer & Wolf, 2010). However, boys showed a slightly better performance in self-efficacy, while girls scored higher on the self-determination scale (Britner, 2008; Glynn et al., 2009). Some other studies found no significant differences with respect to gender in self-efficacy, learning environment stimulation or active learning strategies (Andressa et al., 2016; Chan & Norlizah, 2018). However, girls were more motivated on the science learning values and achievement goal scales (Chan & Norlizah, 2018; King & Ganotice, 2014) as well as on the performance goal (Andressa et al., 2016) and self-efficacy scales in Earth science (Britner, 2008).

The relationship between intelligence and motivation has been considered in previous studies (e.g. Gagné & St Père, 2001; Preckel et al., 2008; Spinath et al., 2006). For example, the study by Spinath and colleagues (2006) revealed that children with higher intelligence are likely to achieve higher academic performance in self-concept, self-efficacy and intrinsic values. A meta-analysis drawing from 74 empirical studies (Kriegbaum et al., 2018) also concluded that intelligence and motivation have a positive relationship and estimated 16.6% of the overall explained variance in school attainment. In addition, Chraif and Dumitru (2015) exposed that students who perform better on IR tests tended to obtain higher points on motivation surveys than their peers. Furthermore, a longitudinal study by Fan and Williams (2010) indicated that there was an aggregate impact of interactive variables, including intrinsic motivation, parental involvement and engagement, on learning mathematics. Overall, it seems that there is a synergistic interaction between cognitive ability and academic motivation in predicting students' achievement in schools.

METHODS OF THE EMPIRICAL STUDIES

A cross-sectional approach can provide acceptable estimates of real developmental process when surveying schoolchildren who develop within slowly changing educational systems. The study samples were selected randomly from 12 public secondary schools in An Giang Province (Vietnam). Table 1 provides the brief of three cross-sectional studies in this project. Students completed instruments in either paper-and-pencil or online administration modes. For online administration, the students accessed the eDia platform and registered for the online instrument with a personal password. For the paper-based assessment, the students were given a test booklet containing the test instruments and an answer sheet.

The background questionnaire was adapted and translated into Vietnamese from the student questionnaire used for PISA 2015 (OECD, 2017). The IR test was adapted from the item bank developed by the Research Group on the Development of Competencies in the University of Szeged (e.g., Csapó, 1997, Korom et al., 2017, Pásztor, 2016). The adapted SR test measures scientific reasoning with the main tasks such as conservation, classification, proportional reasoning, and correlational reasoning. Most of the items were adapted from the original test of Korom et al. (2017) and Lawson (2000)**.** The 24-item test of control of variables in physics test was developed, including three subskills of CVS: identifying controlled experiments, interpreting the outcome of a controlled experiment and understanding the determinacy of confounded experiments. The SMTSL questionnaire was adapted from Tuan, Chin, and Shieh in 2005, involving five subscales: self-efficacy, science learning value, learning strategies, individual learning goals, and learning environment stimulation. Students' motivation and attitude in learning physics questionnaire was adapted and translated into Vietnamese from the student questionnaire in TIMSS 2015 (Hooper et al., 2013; TIMSS, 2015).

Table 1.1. A series of cross-sectional studies.

Timeline	Main aims	Instruments	Samples
September & October 2019	- Validating the instruments - Exploring the trajectories of IR and SM - Examining latent factors predicting individuals' IR	- IR test - SMTSL questionnaire - Background questionnaire	5^{th}, $7t^{h}$, 9^{th} and 11^{th} grades N = 701
October & November 2019	- Exploring developments of IR and SR across cohorts - Investigating the interaction among IR, SR, SM and parental factors in predicting students' STEM achievement	- IR test - SMTSL questionnaire - Background questionnaire	6^{th}, 8^{th}, 10^{th}, and 11^{th} grades N = 733
January & February 2020, 2021	- Developing and validating the CVS test in physics - Investigating development and relationships of CVS and physics motivation in secondary school students - Exploring factors contributing to explaining individual CVS capacity,	- CVS test - SMTSL questionnaire - Background questionnaire	8^{th}, 9^{th}, 10^{th}, 11th, and 12^{th} grades N = 807

Main data analysis strategy includes the factor analysis (Principal component analysis, confirmatory factor analysis, path analysis), Rasch model measurement, differential item functioning, symmetric log-logistic modelling and Bayesian model averaging. The data were analysed with popular statistical programs: R (e.g., R packages such as psych, lavaan and ggplot2), ACER ConQuest and Mplus. Other classical statistical tests such as *t*-test, Pearson's chi-square test, analysis of variance (ANOVA) and Tukey's honestly significant difference (Tukey's HSD) test were used.

RESEARCH AIMS AND EMPIRICAL STUDIES

A set of three cross-sectional studies were proposed to assess reasoning abilities and science motivation of secondary school students. The first empirical study focuses on exploring developmental patterns of IR and SM across grade levels and their association in predicting children's science achievement. Examining relationships of IR and other latent factors of SM offers an in-depth information into reasoning and motivation among individuals, which plays an important role in personalized learning support in enhancing children's academic success in schools.

The second cross-sectional investigation focuses on exploring the developmental curves of inductive reasoning, scientific reasoning abilities, and their roles in predicting students' performance in learning STEM disciplines across grade levels. Students with a better IR ability are likely to earn higher scores on SR tests, and vice versa, IR is a foundational skill in boosting SR. Assessment of the IR and SR and their interaction with background variables can offer a deeper insight into the relationship between cognitive capacities and motivation. It moderately provides evidence for evaluating the extent of the present curricula in terms of optimum cognitive development and learning encouragement in school-age children.

In line with the first two studies, in the third research, we considered examining students' SR and motivation in learning physics. Control of variables strategy (CVS) is a leading component of SR related to domain-general experimentation to evaluate an experimental system to draw valid conclusions. The purposes of this study are to develop, validate the test to measure CVS in physics, and explore development of this ability in secondary school students in Vietnam. The study also outlines basic rationales of assessment of CVS and explores latent predictors of item difficulty as a practical reference for teachers who consider designing assessment-based learning activities in classroom and hands-on tasks in school labs. Across these empirical studies, the effects of mode of administration are also observed by testing measurement equivalence in the test, task and item levels. This might show the initial correspondence of feasibility of applying TBA in Vietnamese context.

Study 1. Exploring inductive reasoning and students' motivation toward science learning across grade levels

The results of CFA showed that the model was a good fit to the values for the cut-off criteria: CFI = .908, TLI = .901, RMSEA = .050 CI (.046, .054) and WRMR = 1.50. In addition, 4-factor model was examined for the IR test with CFA, indicating that it was a well fit to the data, with: CFI = .972, TLI = .970, RMSEA = .028 CI (.023, .032) and WRMR = 1.052. The results of the Rasch analysis indicated a good fit model with the infit for single items (weighted mean squares, MNSQ) ranging from 0.84 to 1.22 (M = 0.98, SD = 0.09). Cronbach's alpha was .88 for the whole test and range of .69 to .81 for individual subtests. The relation between the average item difficulty of 0 logits and the average person proficiency of 1.08 logits scale implied that students' proficiency in IR was higher than the average item difficulty.

In general, students' achievement on the IR test grew gradually throughout the grade levels (Figure 1). The strongest growth occurred from the 5^{th} to 7^{th} grades, and the trend slowed down after the end of lower secondary education (9^{th} grade). The results of ANOVA analysis revealed significant disparities among the grade cohorts on both the subtests and entire test: FS task [$F(3, 697) = 20.78$, $p < .01$], FA task [$F(3, 697) = 25.61$, $p <. 01$], NA task [$F(3, 697) = 36.83$, $p < .01$], NS task [$F(3, 697) = 44.63$, $p <. 01$] and entire test [$F (3, 697) = 51.76$, $p < .01$].

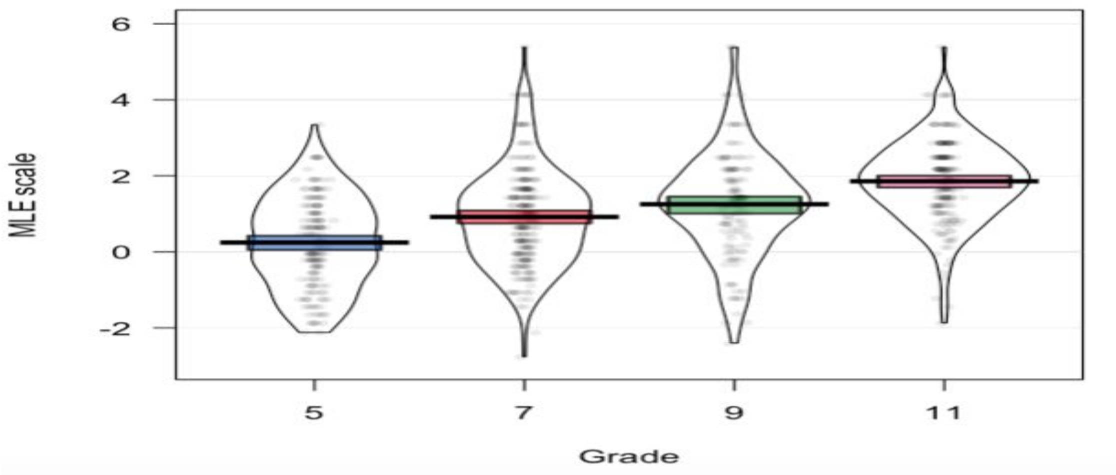

Figure 1. Differences in the students' performance on the IR test regarding grade cohorts.

No significant difference was found on the IR test between boys and girls in the whole sample and in individual cohorts ($p > .05$). This also suggested that students' abilities in each cohort did not differ significantly on the IR test regarding gender. Multiple regression analysis revealed that school grade group, school performance in the previous semester and parents' educational level significantly explained 32.0% of the variance on the IR test, $F (680) = 79.0$, $p < .001$.

For the SMTSL questionnaire, we employed the fit model testing with a bifactor model to the empirical data. The results showed that the model fit is acceptable but not excellent, with cut-off criteria ($\chi^2(139) = 763.3116$, $p < .001$, CFI > .901, RMSEA <. 080, WRMR = 1.627). The internal consistency reliability was generally adequate, with an omega of .87 for the whole items. It was also an acceptable level for the single subscales: SE ($\omega = .69$), AL ($\omega = .80$), LV ($\omega = .56$), AG ($\omega = .74$) and LE ($\omega = .74$).

Generally, students' scores gradually fell grade by grade in science motivation throughout the grade cohorts (Figure 2). Particularly, students' motivation dropped noticeably on the self-efficacy and active learning strategies subscales. In addition, we manipulated the ANOVA analysis to explore the influence of grade levels on the individual subscales and whole questionnaire. Except in the AG subscale, a significant difference was found between the grade cohorts in most subscales: SE [$F(3) = 46.81$, $p<.001$], AL [$F(3) = 32.65$, $p < .001$], LV [$F(3) = 4.27$, $p = .005$], LE [$F(3) = 10.99$, $p < .001$] and SM in general [$F(3) = 5.40$, $p < .001$].

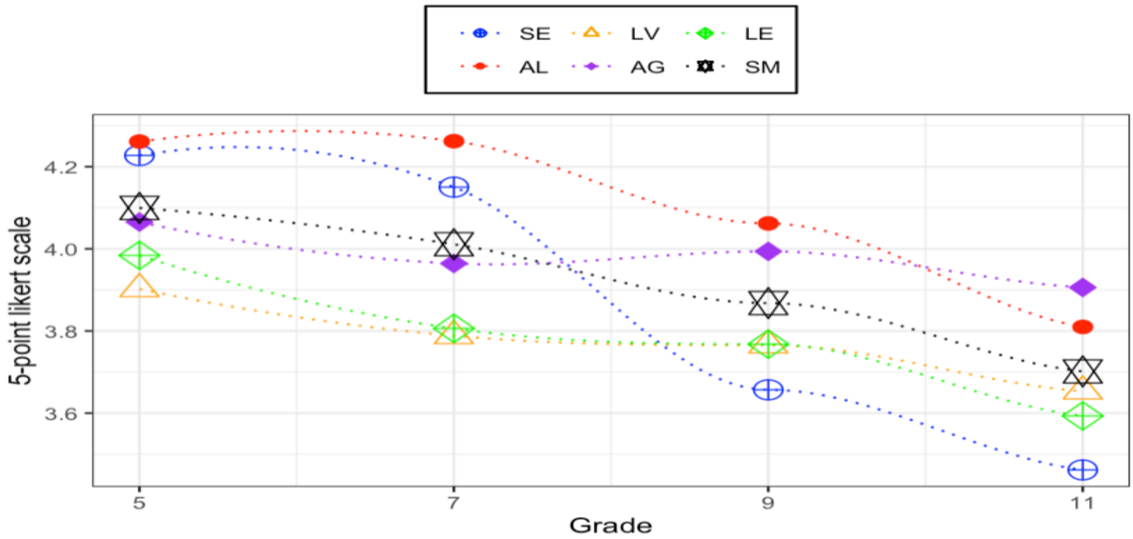

Figure 2. Changes of science motivation on the subscales across grade cohorts.

Note. SE: self-efficacy; AL: active learning strategies; LV: science learning value; AG: achievement goals; LE: learning environment stimulation; SM: science motivation in general.

The *t*-test showed that a significant disparity was only found on the achievement goals subscale, with the girls achieving higher scores (M = 4.05, SD = 0.53) than the boys (M = 3.89, SD = 0.61), t(760.5) = 3.50, $p < .001$. Girls did not differ significantly from boys in most of the subscales and general science motivation in this study. Although a positive correlation was observed between IR and science motivation across grade levels, multi-model Bayesian inference suggested that other factors, such as age, science performance and parental involvement, were better predictors of students' science motivation. Furthermore, a path analysis showed that IR has an indirect effect on science motivation through a science performance variable. The implications for enhancing science motivation are also discussed.

Study 2. Relationship between inductive reasoning, scientific reasoning and science motivation, and their roles in predicting STEM performance

Omega values for the IR test, SR test and SMTSL questionnaire were .82 (Cronbach's alpha = .80), .64 (Cronbach's alpha = .61) and .90 (Cronbach's alpha = .88), respectively, implying that they are acceptable in terms of internal consistency reliability. The unidimensional Rasch model analysis confirmed that all the items on the tests fitted the data quite well. Items of the SMTSL fit well to the data, with the infit values ranging from 0.82 to 1.43, but one item had the infit value higher than 1.3.

For the IR test, the students' mean scores increased remarkably from the 6th (M = 0.73, SD = 1.25) (MLE scale), to the 8th (M = 1.29, SD = 1.29) and on to the 10th grades (M = 1.81, SD = 1.17), before decreasing slightly in the 11th grade (M = 1.59, SD = 1.32). Similar trends were found on the SR test: the students in the older groups performed better than their younger counterparts, except in the 11th grade. In contrast, the students' motivation toward learning science tended to drop slightly across the grade levels. The results of ANOVA analysis showed that there was a significant difference between the grade cohorts on the IR test [$F(3, 722) = 20.53$, $p < .001$], SR test [$F(3, 722) = 29.52$, $p < .001$] and SMTSL questionnaire [$F(3, 722) = 25.1$, $p < .001$].

No significant difference was found between males and females on either the reasoning tests or the SMTSL questionnaire for the entire sample. However, there was a slight difference in cognitive development between males and females. The girls appeared to develop reasoning abilities earlier than the boys, but after the middle school years, the difference remained unchanged or even dropped slightly in the girl group, while boys showed an ongoing improvement until the end of the first year in high school. Unlike reasoning skills, the general trend of student motivation reduced gradually across grade levels.

STEM achievement. The results confirmed that the model (Figure 3) was a good fit to the empirical data, with: $\chi^2(12) = 24.12$, $p = .02$, CFI = .979, TLI = .964, SRMR = .047, RMSEA = .050. The model shows that among the six proposed predictors, only four predictors (inductive reasoning, scientific reasoning, science motivation, and father's education level) have a direct effect and can explain around 25.5% of the STEM achievement variance. Parental involvement in schooling indirectly impacts children's performance through a science motivation variable.

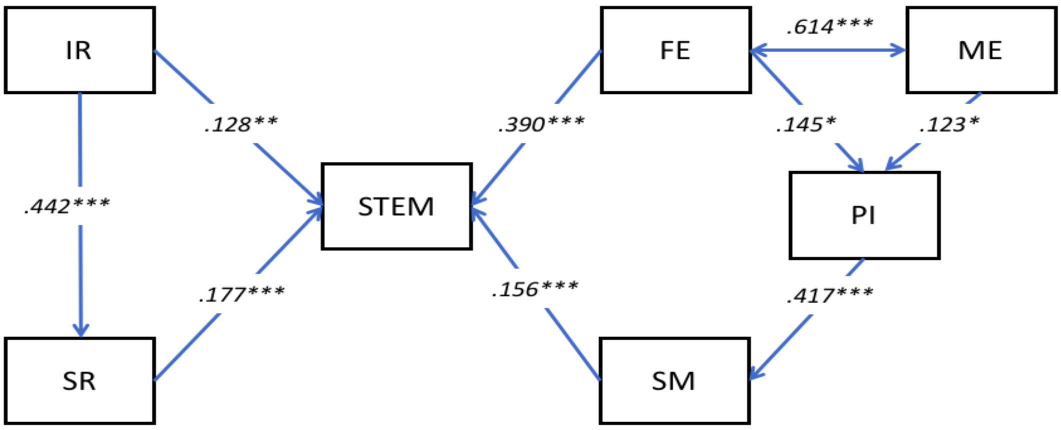

Figure 3. The proposed model for predicting the STEM achievement.

Note. IR: inductive reasoning; SR: scientific reasoning; SM: science motivation; PI: parental involvement in schooling; ME: mother's education level; FE: father's education level.
*p < .05, **p<.01, ***p<.001.

Study 3. Assessing scientific reasoning in control of variables strategy and students' motivation in learning physics in secondary school students

We developed a control of variables strategy in physics (CVSP) test, consisting of 24 multiple-choice items measuring three subskills: identifying controlled experiments (ID), interpreting controlled experiments (IN) and understanding the indeterminacy of confounded experiments (UN). The content of test items involves the knowledge related to basic physics, such as mechanics, heat, thermodynamics, electricity and electromagnetism.

To explore the difference in the internal structure of the test, we fitted unidimensional and three-dimensional models to the dataset. The results showed that a three-dimensional model had a significantly better fit than a unidimensional one. The test results showed an MLE person reliability of .80 (Cronbach's alpha = .81; McDonald's omega = .82), Expected - a posteriori/plausible value reliability of .78, and item reliability of .99. The infit indices ranged from 0.84 to 1.27, and the average value was 0.99 digits (SD = 0.10) in unidimensional model, while three-dimensional model had an infit range from 0.85 to 1.23 and a mean infit of 1.0 (SD = 0.1). This indicated that the test is a reliable construct and statistical fit. Furthermore, we found that item difficulty depended on the subskills but was not affected by physics-related content or the number of independent variables.

The four-parameter symmetric log-logistic equation was applied to describe the developmental process in students' CVS capacities. The logistic curve fit quite well to the empirical data (F = 0.332, p = .565). In general, development of children's CVS in learning physics was statistically significant across grade cohorts, but growth rates were different among particular cohorts. The most rapid growth was flagged at the 10th grade as the point of inflexion (Figure 4). This suggested that the fastest development occurred from the 9th grade to 11th grade ($tan M_{ED50}$ = 2.978).

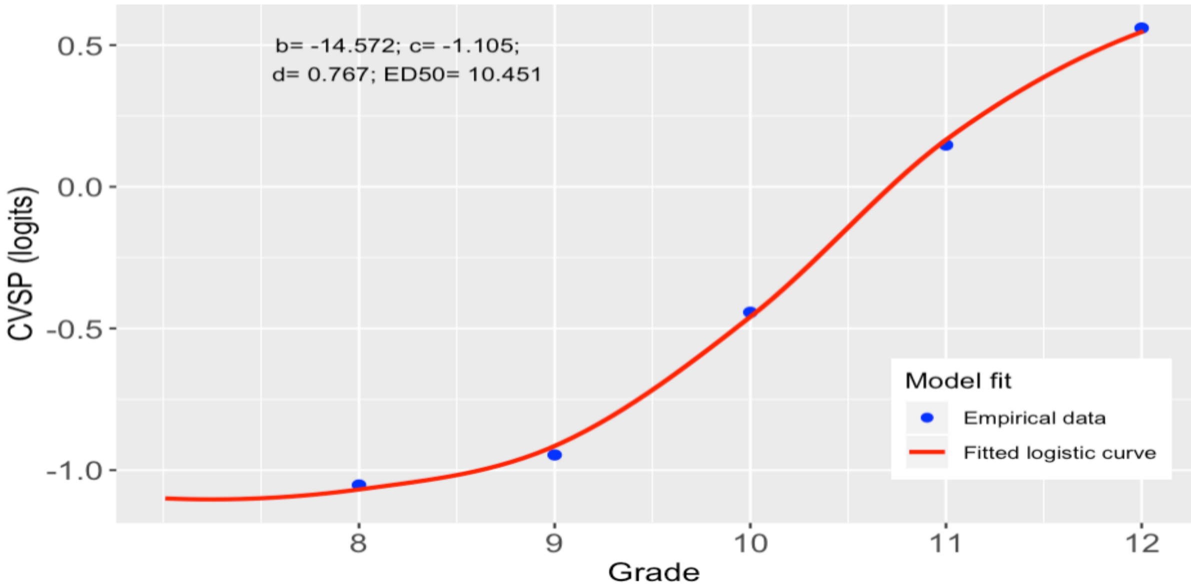

Figure 4. Developmental curve of CVSP in physics.

No significant difference was found on the CVSP test between boys and girls in the whole sample or single grade cohorts. This also suggested that boys did not differ significantly from girls on the CVSP test. However, concerning gender in individual subskills, the results indicated that no gender difference was found on the ID and IN tasks, but males (Mean = −0.97, SD= 1.30) performed significantly better than female (Mean = −1.18, SD= 1.21) on the UN task, t = 2.41, p = .016.

Regarding students' motivation and attitude in learning physics, students' perspectives toward learning physics were positive and quite high in general. The mean scores seemed to be similar among five grade cohorts. Nevertheless, there was a non-significant difference between grade levels on the *Engaged by teaching* scale [F(4, 802) = 2.3, p = .057], but a statistically significant difference was found on the *Like learning* scale among the grade cohorts [F(4, 802) = 7.0, p < .001] and *Confidence in learning* scales [F(4, 802) = 10.7, p < .001].

A path analysis suggested the model presents relationship between understudied variables in contributing to students' CVS proficiency in physics (Figure 5). The model fit well to the empirical data, with χ^2(11) = 23.16, p = .017, CFI = .994, TLI = .984, SRMR = .021, RMSEA = .037 and explained around 44.7% of variance of the CVS capabilities. Grade level demonstrated as the strongest factor affecting directly on the CVS, followed by physics test and LL variables. However, ET and CL do not contribute directly to explaining significantly on the CVS, but CL showed as the main predictor of the physics test.

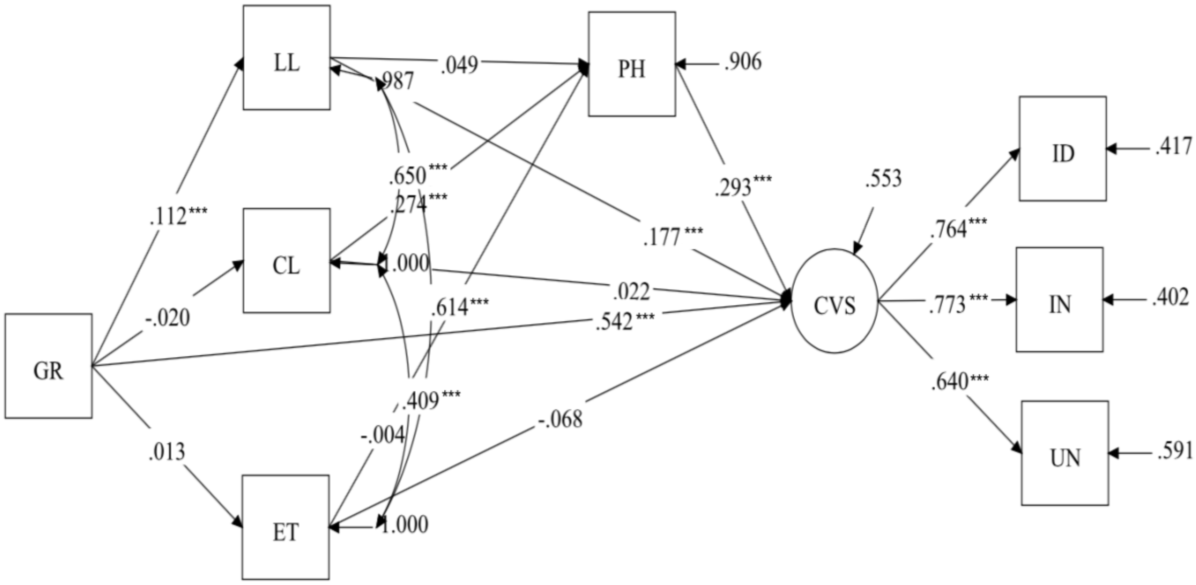

Figure 5. Proposed model predicting students' CVS abilities.

Note. GR: grade level; PH: physics test in previous semester; LL: like learning physics scale; ET: engaging learning in physics; CL: confidence in learning physics.

Study 4. Effects of multimedia on psychometric properties of the cognitive tests: a comparison between technology-based and paper-based assessment

The main aim of the investigations is to examine the effects of technology on administration modes on the cognitive reasoning tests across three empirical studies at the item, task, and test levels. Regarding the average total score, under teachers' supervision, the results seemed to be more supportive for the paper-and-pencil version because the students performed better than their peers did via online assessment in the same test. On the other hand, without supervision, the results seemed to favour the online format as the students who took online testing performed better than their friends who participated in the paper-based test version.

Interestingly, across delivery conditions, the results demonstrated that either with or without supervision, Rasch model measurement indicated that the online assessment had a better fit to the empirical data than the paper one. The results of the measurement invariance in DIF analysis with Angoff's Delta approach exposed the internal validity of the tests seemed acceptably comparable across the two modes of administration. The reliability and validity of the two tests were equivalent regardless of modes of media delivery. The study provided evidence of linking between test modality and the materials constructing the test items. It further illuminates how TBA can be more beneficial when items are composed of visually rich materials.

GENERAL DISCUSSION

A systematic review revealed that among several forms of reasoning, IR and SR gain considerable interests in educational contexts. Though the types of reasoning tasks appeared unchanged over twenty-three years, they grew more diverse and gradually evolved from PP to online formats available in smartphones, tablets and wearable devices. The literature review also discovered that reasoning skills are closely associated with several variables, such as age, discipline performance, parental factors and problem-solving skills, while gender differences may depend on particular cultures.

The empirical studies confirmed the quality of the adapted tests and questionnaires by undergoing the investigations of validity and reliability across grade levels. We also developed a new test to measure scientific reasoning in CVS in physics for secondary school students in Vietnamese context. The initial analysis showed that the test served as a reliable instrument tool to assess the CVS capacity in Vietnamese students. Equivalences are often expected between different delivery modalities and gender; and thus, our studies also used multifaceted methods to observe measurement invariance of the tests and questionnaires. The feasibility of transferring from PP to TBA was examined through the empirical studies. The findings suggested that the results from the online formats showed a better fit to the data than those from PP ones.

Students' performance on the reasoning tests increased grade by grade during secondary school education years. The developmental curves of IR and SR in students demonstrated similar results across the grade cohorts, in which the most rapid development was observed between the age group of 12 - 14 years (6^{th} - 8^{th} grades), but the growth rate apparently slowed down after those years. Meanwhile, the development of children's CVS fitted well to the symmetric log-logistic curve, and the fastest growth was detected at the age range of the 14 - 16 years (the 10^{th} grade). The changing patterns of SM in children seemed fluctuated and even somewhat reduced in some motivational components across grade cohorts. Specially, students' scores on self-efficacy and active learning strategy scales reduced drastically in the older groups. The results showed that general reasoning proficiencies in children develop most rapidly at the middle school level, suggesting that lower secondary school is the most appropriate time to nurture thinking skills in the school curricula. Thus, parents and teachers should be aware of the golden opportunity to boost children's reasoning skills in those years by teaching school subjects and practising daily activities which can enhance their reasoning skills.

The findings revealed that cognitive abilities, motivation, and parental factors are the important predictors of children's STEM achievement. Parents' education levels and parental involvement in schooling impacted meaningfully on their kids' performance in schools. Ample parental support and engagement as well as interest in school activities play a key role in children's motivation in learning and engaging in school activities. Notably, most of the students in all grade cohorts reported lowest scores in the learning environment stimulation scale. Similarly, students' confidence in learning physics were quite low indicated by the lower mean score, compared with other scales. Correspondingly, both schools and families should be concerned about how to inspire children's motivation to learn science. Teachers, school leaders and school psychologists should be aware of these trends to find more supportive facilities to improve learning environments and enhance children's confidence in learning science at schools.

There are some limitations in the investigations. The theoretical groundings are endeavoured based on the popular concepts of reasoning and motivation which may be a controversial topic in the psychological field. It is the first time that these test instruments were conducted in Vietnamese context, so some items need to be revised in order to improve validity and reliability for next-generation versions. Other concern may also arise from the study sample which was conducted in An Giang province (Vietnam) with cross-sectional investigations. More large-scale assessment and longitudinal approach need to be managed to weigh the success of the current curricula.

REFERENCES

Adey, P., & Csapó, B. (2012). Developing and assessing scientific reasoning. In B. Csapó & G. Szabó (Eds.), *Framework for diagnostic assessment of science* (pp. 17–53). Nemzeti Tankönyvkiadó.

Andressa, H., Mavrikaki, E., & Dermitzaki, I. (2016). Adaptation of the Students' Motivation Towards Science Learning Questionnaire To Measure Greek Students' Motivation Towards Biology Learning. *International Journal Of Biology Education*, *4*(2). https://doi.org/10.20876/ijobed.56334

Blum, D., Holling, H., Galibert, M. S., & Forthmann, B. (2016). Task difficulty prediction of figural analogies. *Intelligence*, *56*, 72–81. https://doi.org/10.1016/j.intell.2016.03.001

Britner, S. L. (2008). Motivation in high school science students: A comparison of gender differences in life, physical, and earth science classes. *Journal of Research in Science Teaching*, *45*(8), 955–970. https://doi.org/10.1002/tea.20249

Chan, Y. L., & Norlizah, C. . (2018). Students' motivation towards science learning and students' science achievement. *International Journal of Academic Research in Progressive Education and Development*, *6*(4), 174–189. https://doi.org/10.6007/IJARPED/v6-i4/3716

Chraif, M., & Dumitru, D. (2015). Differences between motivation from competition and motivation from individual goals under the influence of inductive reasoning. *Procedia - Social and Behavioral Sciences*, *187*(2015), 745–751. https://doi.org/10.1016/j.sbspro.2015.03.157

Csapó, B. (1997). The development of inductive reasoning: Cross-sectional assessments in an educational context. *International Journal of Behavioral Development*, *20*(4), 609–626. https://doi.org/10.1080/016502597385081

Díaz-Morales, J. F., & Escribano, C. (2013). Predicting school achievement: The role of inductive reasoning, sleep length and morningness-eveningness. *Personality and Individual Differences*, *55*(2), 106–111. https://doi.org/10.1016/j.paid.2013.02.011

Ding, L. (2018). Progression trend of scientific reasoning from elementary school to university: a large-scale cross-grade survey among Chinese students. *International Journal of Science and Mathematics Education*, *16*(8), 1479–1498. https://doi.org/10.1007/s10763-017-9844-0

Du, N. N. (2015). *Factors influencing teaching for critical thinking in Vietnamese lower secondary schools: a mixed method study focussed on history (Unpublished doctoral dissertation)*. Newcastle University.

Fan, W., & Williams, C. M. (2010). The effects of parental involvement on students' academic self-efficacy, engagement and intrinsic motivation. *Educational Psychology*, *30*(1), 53–74. https://doi.org/10.1080/01443410903353302

Gagné, F., & St Père, F. (2001). When IQ is controlled, does motivation still predict achievement? *Intelligence*, *30*(1), 71–100. https://doi.org/10.1016/S0160-2896(01)00068-X

Ganzach, Y. (2000). Parents' education, cognitive ability, educational expectations and educational attainment: Interactive effects. *British Journal of Educational Psychology*, *70*(3), 419–441. https://doi.org/10.1348/000709900158218

Garcia, T., & Pintrich, P. R. (1995). The role of possible selves in adolescents' perceived competence and self-regulation. *Annual Meeting of the American Research Association*.

Glynn, S. M., Taasoobshirazi, G., & Brickman, P. (2009). Science motivation questionnaire: Construct validation with nonscience majors. *Journal of Research in Science Teaching*, *46*(2), 127–146. https://doi.org/10.1002/tea.20267

Hamers, J. H. M., De Koning, E., & Sijtsma, K. (1998). Inductive reasoning in third grade: Intervention promises and constraints. *Contemporary Educational Psychology*, *23*(2), 132–148. https://doi.org/10.1006/ceps.1998.0966

Han, J. (2013). *Scientific reasoning: research, development, and assessment (Unpublished doctoral dissertation)*. The Ohio State University.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hooper, M., Mullis, I. V. S., & Martin, M. O. (2013). TIMSS 2015 context questionnaire framework. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2015 Assessment Frameworks* (pp. 61–82). TIMSS & PIRLS International Study Center.

Hwang, M. H., Choi, H. C., Lee, A., Culver, J. D., & Hutchison, B. (2016). The relationship between self-efficacy and academic achievement: A 5-year panel analysis. *The Asia-Pacific Education Researcher*, *25*(1), 89–98. https://doi.org/10.1007/s40299-015-0236-3

Jeotee, K. (2012). *Reasoning skills, problem solving ability and academic ability: implications for study programme and career choice in the context of higher education in Thailand (Unpublished doctoral dissertation)* [Durham University]. http://etheses.dur.ac.uk/3380

Kambeyo, L. (2018). *Assessing Namibian students ' abilities in scientific reasoning, scientific inquiry and inductive reasoning skills (Unpublished doctoral dissertation)*. University of Szeged.

Kambeyo, L., & Wu, H. (2018). Online assessment of students' inductive reasoning skills abilities in Oshana region , Namibia. *International Journal of Educational Sciences*, *21*, 1–12. https://doi.org/11.258359/KRE-86

King, R. B., & Ganotice, F. A. (2014). What's happening to our boys? A personal investment analysis of gender differences in student motivation. *Asia-Pacific Education Researcher*, *23*(1), 151–157. https://doi.org/10.1007/s40299-013-0127-4

Kinshuk, Lin, T., & Mcnab, P. (2006). Cognitive trait modelling: The case of inductive reasoning ability. *Innovations in Education and Teaching International*, *43*(2), 151–161. https://doi.org/10.1080/14703290600650442

Köksal-tuncer, Ö., & Sodian, B. (2018). The development of scientific reasoning : Hypothesis testing and argumentation from evidence in young children. *Cognitive Development*, *48*(2018), 135–145. https://doi.org/10.1016/j.cogdev.2018.06.011

Korom, E., B. Németh, M., Pásztor, A., & Csapó, B. (2017). Relationship between scientific and inductive reasoning in grades 5 and 7. *Paper Presented at the 17th Biennial Conference of the European Association for Research on Learning and Instruction (EARLI)*.

Kriegbaum, K., Becker, N., & Spinath, B. (2018). The relative importance of intelligence and motivation as predictors of school achievement: A meta-analysis. *Educational Research Review*, *25*(October), 120–148. https://doi.org/10.1016/j.edurev.2018.10.001

Kwon, Y.-J., & Lawson, A. E. (2000). Linking brain drowth with the development of scientific reasoning ability and conceptual change during adolescence. *Journal of Research in Science Teaching*, *37*(1), 44–62. https://doi.org/10.1002/(SICI)1098-2736(200001)37:1<44::AID-TEA4>3.0.CO;2-J

Kyllonen, P., Hartman, R., Sprenger, A., Weeks, J., Bertling, M., McGrew, K., Kriz, S., Bertling, J., Fife, J., & Stankov, L. (2019). General fluid/inductive reasoning battery for a high-ability population. *Behavior Research Methods*, *51*(2), 507–522. https://doi.org/10.3758/s13428-018-1098-4

Lawson, A. (2000). Classroom test of scientific reasoning. In *Revised Edition: August 2000 by Anton E. Lawson, Arizona State University. Based on: Lawson, A.E. 1978. Development and validation of the classroom test of formal reasoning. Journal of Research in Science Teaching, 15(1): 11-24.* http://www.public.asu.edu/~anton1/AssessArticles/Assessments/Mathematics Assessments/Scientific Reasoning Test.pdf
Lawson, A. (2009). Basic inferences of scientific reasoning, argumentation, and discovery. *Science Education*, *94*(2), 336–364. https://doi.org/10.1002/sce.20357
Leighton, J. P., & Sternberg, R. J. (2004). The nature of reasoning. In *The Nature of Reasoning* (pp. 3–11). Cambridge University Press. https://doi.org/10.1161/CIRCRESAHA.116.305012
Mayer, D., Sodian, B., Koerber, S., & Schwippert, K. (2014). Scientific reasoning in elementary school children: Assessment and relations with cognitive abilities. *Learning and Instruction*, *29*, 43–55. https://doi.org/10.1016/j.learninstruc.2013.07.005
Molnár, G. (2011). Playful fostering of 6- to 8-year-old students' inductive reasoning. *Thinking Skills and Creativity*, *6*(2011), 91–99. https://doi.org/10.1016/j.tsc.2011.05.002
Molnár, G., Greiff, S., & Csapó, B. (2013). Inductive reasoning, domain specific and complex problem solving: Relations and development. *Thinking Skills and Creativity*, *9*, 35–45. https://doi.org/10.1016/j.tsc.2013.03.002
Muniz, M., Seabra, A. G., & Primi, R. (2012). Validity and reliability of the inductive reasoning test for children - IRTC. *Psicologia: Reflexão e Crítica*, *25*(2), 275–285. https://doi.org/10.1590/s0102-79722012000200009
Nikolov, M., & Csapó, B. (2018). The relationships between 8th graders' L1 and L2 reading skills, inductive reasoning and socio-economic status in early English and German as a foreign language programs. *System*, *73*, 48–57. https://doi.org/10.1016/j.system.2017.11.001
Nunes, T., & Csapó, B. (2011). Developing and assessing mathematical reasoning. In Benő Csapó & M. Szendrei (Eds.), *Framework for diagnostic assessment of mathematics* (pp. 15–76). Nemzeti Tankönyvkiadó.
OECD. (2017). *PISA 2015 results students' well-being: Vol. III*. OECD. https://doi.org/10.1787/9789264273856-en
Osborne, J. (2013). The 21st century challenge for science education: Assessing scientific reasoning. *Thinking Skills and Creativity*, *10*, 265–279. https://doi.org/10.1016/j.tsc.2013.07.006
Pásztor, A. (2016). *Technology-based assessment and development of inductive reasoning*. Ph.D. thesis. Doctoral School of Education, University of Szeged. doi:10.14232/phd.3191
Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education: Theory, research, and applications*. Prentice Hall.
Piraksa, C., Srisawasdi, N., & Koul, R. (2014). Effect of Gender on Student's Scientific Reasoning Ability: A Case Study in Thailand. *Procedia - Social and Behavioral Sciences*, *116*, 486–491. https://doi.org/10.1016/j.sbspro.2014.01.245
Preckel, F., Goetz, T., Pekrun, R., & Kleine, M. (2008). Gender Differences in Gifted and Average-Ability Students. *Gifted Child Quarterly*, *52*(2), 146–159. https://doi.org/10.1177/0016986208315834
Salihu, L., Aro, M., & Räsänen, P. (2018). Children with learning difficulties in mathematics: Relating mathematics skills and reading comprehension. *Issues in Educational Research*, *28*(4), 1024–1038.
Schweizer, F., Wüstenberg, S., & Greiff, S. (2013). Validity of the MicroDYN approach: Complex problem solving predicts school grades beyond working memory capacity.

Learning and Individual Differences, *24*, 42–52. https://doi.org/10.1016/j.lindif.2012.12.011
Soodmand Afshar, H., Rahimi, A., & Rahimi, M. (2014). Instrumental motivation, critical thinking, autonomy and academic achievement of iranian EFL learners. *Issues in Educational Research, 24*(3), 281–298.
Spinath, B., Spinath, F. M., Harlaar, N., & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence, 34*(4), 363–374. https://doi.org/10.1016/j.intell.2005.11.004
Sternberg, R. J., & Sternberg, K. (2012). *Cognitive Psychology*. Cengage Learning products. https://doi.org/10.1039/ft9918702861
Tairab, H. H. (2015). Assessing students' understanding of control of variables across three grade levels and gender. *International Education Studies*, *9*(1), 44–54. https://doi.org/10.5539/ies.v9n1p44
Tekkaya, C., & Yenilmez, A. (2006). Relationships among measures of learning orientation, reasoning ability, and conceptual understanding of photosynthesis and respiration in plants for grade 8 males and females. *Journal of Elementary Science Education*, *18*(1), 1–14. https://doi.org/10.1007/BF03170650
Thuneberg, H., Hautamäki, J., & Hotulainen, R. (2015). Scientific reasoning, school achievement and gender: a Multilevel study of between and within school effects in Finland. *Scandinavian Journal of Educational Research, 59*(3), 337–356. https://doi.org/10.1080/00313831.2014.904426
TIMSS. (2015). *Student questionnaire (separate Science Subjects)*.
Tuan, H. L., Chin, C. C., & Shieh, S. H. (2005). The development of a questionnaire to measure students' motivation towards science learning. *International Journal of Science Education*, *27*(6), 639–654. https://doi.org/10.1080/0950069042000323737
Venville, G., & Oliver, M. (2015). The impact of a cognitive acceleration programme in science on students in an academically selective high school. *Thinking Skills and Creativity*, *15*(2015), 48–60. https://doi.org/10.1016/j.tsc.2014.11.004
Zeyer, A. (2010). Motivation to learn science and cognitive style. *Eurasia Journal of Mathematics, Science and Technology Education*, *6*(2), 123–130. https://doi.org/10.12973/ejmste/75233
Zeyer, A., & Wolf, S. (2010). Is there a relationship between brain type, sex and motivation to learn science? *International Journal of Science Education*, *32*(16), 2217–2233. https://doi.org/10.1080/09500690903585184
Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, *20*(1), 99–149. https://doi.org/10.1006/drev.1999.0497
Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*(2), 172–223. https://doi.org/10.1016/j.dr.2006.12.001

RELEVANT PUBLICATIONS

Journal articles

Van Vo, D., & Csapó, B. (*in press*). Measuring inductive reasoning in school contexts: A review of instruments and predictors. *International Journal of Innovation and Learning*.

Van Vo, D., & Csapó, B. (2021). Development of scientific reasoning test measuring control of variables strategy in physics for high school students: evidence of validity and latent predictors of item difficulty. *International Journal of Science Education*. https://doi.org/10.1080/09500693.2021.1957515

Van Vo, D., & Csapó, B. (2021). Exploring students' science motivation across grade levels and the role of inductive reasoning in science motivation. *European journal of Psychology of Education*. https://doi.org/10.1007/s10212-021-00568-8

Van Vo, D., & Csapó, B. (2020). Development of inductive reasoning in students across school grade levels. *Thinking Skills and Creativity*, *37*(2020), 100699. https://doi.org/10.1016/j.tsc.2020.100699

Conference papers

Van Vo, D. (2021). Interactions of Reasoning Abilities and Science Motivation with Parent Involvement Variables in Predicting the STEM Achievement. In: Molnár, Gyöngyvér; Tóth, Edit (ed.): *The answers of education to the challenges of the future: XXI. ONK*. National Educational Science Conference. November 18-20, 2021. Szeged, Hungary: Institute of Education, University of Szeged, pp. 218-218.

Van Vo, D., & Csapó, B. (2021). A Comparison between Technology-Based and Paper-Based Assessment on the Scientific Reasoning Test in Control of Variables Strategy in Physics. In: Molnár, Gyöngyvér; Tóth, Edit (ed.): *The answers of education to the challenges of the future: XXI. ONK*. National Educational Science Conference. November 18-20, 2021. Szeged, Hungary: Institute of Education, University of Szeged, pp. 167-167.

Van Vo, D.(2021). Development of inductive reasoning and its relationship with science performance in Vietnam. Abtract book: *The 14th Training and Practice International Conference on Educational Science*. Kaposvár, Hungary: Kaposvár University Faculty of Pedagogy, pp. 87-87.

Van Vo, D. (2021). Exploring the patterns of inductive reasoning, scientific reasoning and science motivation in Vietnamese students across grade levels. Abstract book: *In EARLI 2021: Education and Citizenship: Learning and Instruction and the Shaping of Futures*. Online, pp. 230-230.

Van Vo, D. (2021). Assessing control of variables strategy in physics of high school students. Abstract book: *In EARLI 2021: Education and Citizenship: Learning and Instruction and the Shaping of Futures*. Online, pp. 305-305.

Van Vo, D. (2020). Motivation toward science learning and inductive reasoning ability of secondary students in Vietnam: developmental changes and relationships. Abstract book: *VI. Ipszilon Konferencia*. Budapest, Hungary: ELTE PPK, pp. 48-49.

Van Vo, D. (2020). Assessment of motivation toward science learning and scientific reasoning of students in Vietnam: developments and relationships. Abtract book: *The*

13th Training and Practice International Conference on Educational Science. Kaposvár, Hungary: Kaposvár University Faculty of Pedagogy, pp. 111-111.

Van Vo, D. (2019). Technology-based assessment of reasoning skills: A literature review. In: Varga, Aranka; Andl, Helga; Molnár-Kovács, Zsófia (eds.) Neveléstudomány – Horizontok és dialógusok. Absztraktkötet. : *XIX. ONK*. Pécs, 2019. november 7-9. Pécs, Hungary : Pécsi Tudományegyetem Bölcsészettudományi Kar Neveléstudományi Intézet, pp. 383-383.

Van Vo, D. (2019). *Technology-based assessment in the educational setting.* In: Molnár, Edit Katalin; Dancs, Katinka (eds.) *CEA 2019: 17th Conference on Educational Assessment* . Programme and Abstracts. Szeged, Hungary: Szegedi Tudományegyetem, pp. 98-98.