The Four-Component Instructional Design Model (4C/ID): A meta-analysis on use and effect

Abstract

The four-component instructional design model (4C/ID) has been increasingly used in face-to-face and online learning environments. We present a meta-analysis on the use and effect of educational programs developed with the 4C/ID model on performance, after more than 20 years of its application and research in different academic areas and technical training. We performed the meta-analysis through the combination of the effect sizes of the studies using Cohen's d. The combination of the studies suggests that the use of educational programs developed with 4C/ID has a high impact on performance ($d = 0.79$), regardless of the academic area, the design of the study and the outcome (knowledge and complex skills). The grade under study was a significant moderator on the effect, showing that the higher-education level is more suitable for application of the 4C/ID model. Our results suggest that the use of the 4C/ID model should be prioritized as instructional model in college and university learning environments.

Keywords: 4C/ID; instructional design; meta-analysis; complex skills; performance

Declarations

Funding

This work was partially supported by Fundação para a Ciência e a Tecnologia, I.P. (FCT) [ISTAR Projects: UIDB/04466/2020 and UIDP/04466/2020]

Conflicts of interest

The authors declare that there is no conflict of interest.

Availability of data and material

Data will be available upon request.

Code availability (software application or custom code)

Not applicable.

Introduction

Instructional Design can be defined as the domain of human activity that concerns learning and performance improvement (Merril 2002). It aims to optimize learning rather than describe or explain it, and therefore has a prescriptive and normative orientation (Bruner 1966).

Reiser (2001) highlights that instructional design is more than the use of certain resources and technologies; its purpose is to improve performance. Recently, Reigeluth, Beatty and Myers (2017) gave us a vision of changes taking place in instruction theory, arising from the technological society from the Information Age, contributing to the shift from the teacher-centered to learner-centered models. These changes have had repercussions in several domains associated with the instructional design: instructional management, assessment and even curriculum. According to these authors, the development of the theory of instruction arose during the 1980s. At that time, systemic or process models predominated, such as the ADDIE model (Analysis, Design, Development, Implementation, Evaluation), still used nowadays. During the 1990s, based on the urgency of a new paradigm and its different theories, research was concerned on centering the instruction on the learner rather than on the teacher (Reigeluth et al. 2017). From 2009, academics tried to find a common language and knowledge in this field, without replacing the different models and instructional theories that emerged mainly during the 1990s.

Each instructional theory is associated with a scientific theory of cognitive development or human learning, which explains how we acquire and transfer knowledge

(Bruner 1966). Although instructional design as a disciplinary field emerged with Skinner's Programmed Instruction in the early fifties (Skinner 1954), today there are three major approaches to instructional design: instructivism, rooted in the behaviorist theories of learning; constructivism, rooted in cognitive developmental psychology; and cognitivism, rooted in information process theories or cognitive psychology.

Cognitive models base their proposals on experimental research that is developed within the information processing psychology or cognitive psychology, where memory is the main mechanism that allows human beings to learn (e.g. Anderson 1983, 1993; Baddeley 1997; Nisser 1967). They value the processes of knowledge construction that follow the way memory works: a working memory with limited capacity (Baddeley 1997) which processes the different stimuli that arrive to the human being through the sensorial systems transforming them into mental representations. This process makes the connection between this new knowledge and previously acquired knowledge organized in long-term memory in mental schemas. In human cognition framework, learning means acquiring new mental schemas, integrating others already acquired into broader schemas and automating other schemas (eg. Anderson 1983, 1993; Blessing and Anderson 1996).

One of the cognitive instructional models that is currently used in many scientific areas as a dimension of powerful learning environments is van Merriënboer's Four-Component Instructional Design Model (4C/ID) (van Merriënboer and Paas 2003). According to its authors, the 4C/ID model is an instructional design model mainly suitable for teaching complex learning (Frerejean, van Merriënboer, Kirschner, Roex, Aertgeerts, and Marcellis 2019). It is based on the cognitive psychology or information processing approach to cognition, where memory plays a central role in the process of learning. The 4C/ID model also advocates a task-centered instructional design that the traditional cognitive models, as the conditions-based model of Gagné, do not share. According to

Merrill (2002), this model is "perhaps, the most comprehensive recent model of instructional design that is problem-centered" (p. 56), involving all of the first principles of instruction.

The 4C/ID model allows the development of powerful learning environments that integrate some of the characteristics of categories of learning environments proposed by Bransford, Brown and Cocking (2000): student-centered, knowledge-centered, assessment-centered and community-centered. For Bransford, Brown and Cocking (2000), learning environments that manage to bring all these perspectives together are the ones that are most effective in learning, both inside and outside school.

The 4C/ID Model

The 4C/ID model was developed by van Merriënboer and his colleagues in the early 1990s. Its main objective is to support the teaching of complex learning, requiring the integration of knowledge, complex skills and attitudes to solve real-world problems (van Merriënboer and Kirschner 2001; van Merriënboer and de Croock 1992). This model covers the current knowledge about human cognitive architecture, the limitations of Working Memory, and a set of principles proposed by the cognitive theory of multimedia learning (eg. Mayer 2014) and the cognitive load theory (eg. Sweller, Ayres, and Kalyuga 2011; Sweller, van Merriënboer and Paas 2019), aiming to promote the acquisition and transfer of knowledge (van Merriënboer and Kester 2004).

Van Merriënboer, Jelsma, and Paas (1992) made the first presentation to the scientific community of the 4C/ID model responsible to train complex skills. They described the different stages of implementing the instructional model for training complex skills (design, presentation of information to support learning and training). Their work also presented a practical application of the model to the development of professional skills in distinct areas, such as engineering, programming and statistics. The authors predicted that the application of the 4C/ID model produces more reflective knowledge and better performance in learning transfer (van Merriënboer et al. 1992). Following this work, van Merriënboer, Clark and de Croock (2002) described a framework for the application of the 4C/ID model for complex competence training, suggesting the structure of the training plan and the associated instructional methods.

The 4C/ID model is composed by four interconnected components (Figure 1): (i) *learning tasks*, which are the central feature of the model, being preferentially based on real-world problems; (ii) *supportive information,* that helps the learner to solve more efficiently the problems defined in the learning tasks, connecting previously acquired knowledge to the newly presented knowledge, (c) *procedural information,* which provides information about how the routine aspects of the tasks should be performed; this information is organized into small segments of information presented at the exact moment they are needed – a just-in-time information; and (d) *part-task practice* that allows the learner to train routine skills that are part of the *learning tasks* and should become fully automated.

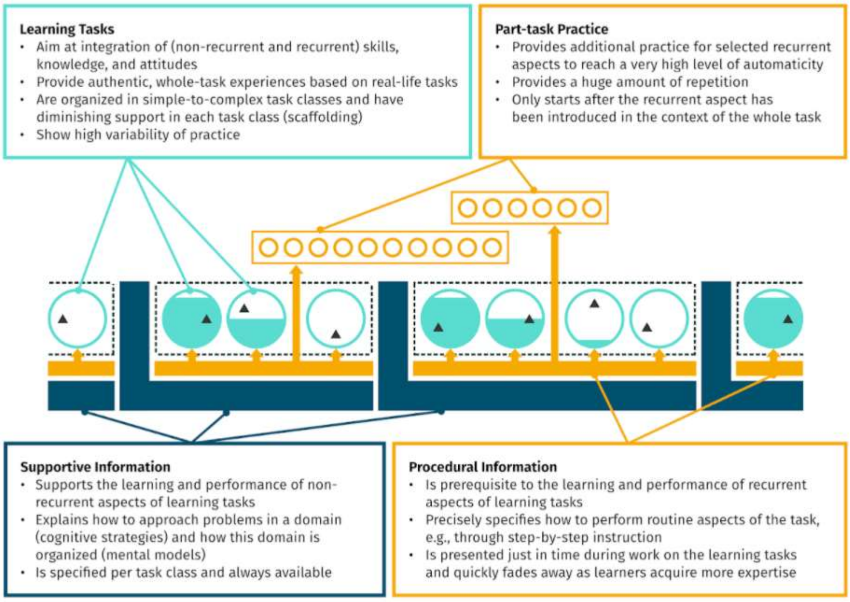

Figure 1. 4C/ID model (reprinted from 4CID.org 2020)

In the 4C/ID model, it is critical to carefully manage learners' cognitive load, because of its focus on whole-task learning tasks. Due to the task complexity of this kind of learning tasks it is crucial to accurately manage the cognitive load imposed on the learners.

In order to manage these types of cognitive load, the 4C/ID model suggests the following specific strategies in terms of task sequencing and information presentation:

a) *Sequence learning tasks classes from simple to complex*. Task complexity could compromise the learning process because of the excessive cognitive overload for the learners. Thus, in order to fix this problem the first task class should be the simplest version of the whole tasks the learners encounter in the work context. The final task class should be the most complex, including the real-world tasks;

b) *Sequence learning tasks with decreasing learner support (fading from high to no support)*. Another way to prevent learners' cognitive overload is the use of various types of learning tasks ranging from high built-in-support, via tasks with an intermediate level of support, to conventional tasks without support (e.g., worked examples and completion tasks);

c) *Sequence learning tasks in a variable order (high contextual interference)*. Research findings indicate that high variability of practice affects the development of schemata and promotes subsequent learning transfer;

d) *Present supportive information before learners start working on the learning tasks, and make it accessible to learners during the practice*, because this information involves mental models (e.g., what is this?, how is this organized, how does this work?) and cognitive strategies, with high complexity. So, performing practice and studying supportive information simultaneously may cause cognitive overload (Kester, Kirschner, van Merriënboer and Baumer 2001);

e) *Present just-in-time information when learners need it*, to reduce ineffective cognitive load caused by temporal split attention (Mayer and Sims 1994).

Research on the 4C/ID model.

The research carried out under the 4C/ID model has been developed in two main areas: (i) one focused on improving the applicability conditions of the learning environment that integrates the model, for example, the format and timing of the presentation of supportive information, how to organize the learning tasks or how to sequence the task classes (e.g. Paas 1992; Nadolski, Kirschner and van Merriënboer 2004, 2006); (ii) the other is related to the efficiency of the model on the acquisition of knowledge and its transfer to other tasks, more or less distant from the previous ones (e.g. van Es and Jeuring 2017; Martínez-Mediano and Losada 2017); or even combining performance (acquisition and transfer) with the perceived mental effort (e.g. Sarfo and Elen 2007; Melo and Miranda 2015).

We carried out a meta-analysis over the studies included in (ii), aiming to understand the effects of the whole model on the acquisition and transfer of knowledge.

Method

To understand the effects of the 4C/ID model on learning, we developed a systematic review following the PRISMA statement (Moher, Liberati, Tetzlaff and Altman 2009).

Information Sources and Search

We searched for the keywords ("4C" and "instructional design") to include the abbreviations 4C/ID, 4CID and 4C-ID presented in the literature, or the full name of the model as search term ("four-component instructional design") in topic, abstract or title. We only searched for articles from 1992 since this was the year of the first publication of the model (van Merriënboer, Jelsma and Paas 1992). We performed all database searches in the first week of May 2020.

The selected databases were the major ones that covered high quality studies in the field, namely, Web of Science (WoS) of Clarivate Analytics, including all databases in the WoS Core Collection, Scopus, ERIC, DOAJ and IEEEXplore. We also looked for PhD theses in DigiNole using the same search keywords.

Eligibility Criteria

After the search phase, the studies that follow the next criteria were included:

i. The study includes the use of instructional materials based on the whole 4C/ID model. Studies that use only a segment of the 4C/ID model were excluded;
ii. The study presents quantitative empirical results;
iii. Results provide or allow the calculation of effect size;
iv. Results are not provided from self-assessment;
v. The study measures knowledge or complex skills.

Figure 2 shows the flow of information through this systematic review, based on Moher's et al. (2009) PRISMA diagram.

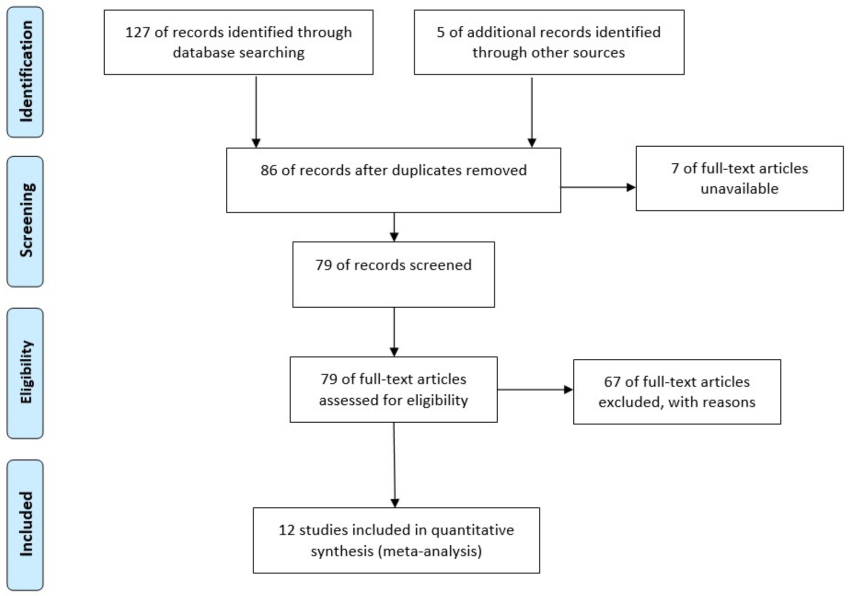

Figure 2. PRISMA flow diagram through the 4C/ID model meta-analysis.

From the 127 records identified in searching databases (Scopus 50, DOAJ 4, IEEEXplore 7, WoS 54, DigiNole 3), 44 were repeated through databases and seven (mainly book chapters) were unavailable.

We also searched on the127 records' references identified in search databases. From the records' references we got two additional records that matched the inclusion criteria.

Statistical analysis

We used Cohen's d to calculate the effect size of the select studies with the support of Lenhard and Lenhard's application (2016). The combination of the studies with effect sizes allows us to interpret studies through a single measure (Borenstein, Hedges, Higgins and Rothstein 2009) and also to explain the characteristics that are responsible for the differences in the results (Coe 2002).

The Cohen's d represents the estimates of the quantity of subjects of the experimental group that are expected to exceed the medium value of the control group when represented by a proportion (Conboy 2003). This measure has the advantage of using an unbiased estimative to true population variability, whether the null hypothesis is true or not. Also, it is less biased than Hedges' g (Conboy 2003).

Additionally, we use a p-curve method to investigate whether the significant results present in the literature can be explained by phenomena such as p-hacking and publication bias, or if they signify real evidentiary value (Simonsohn, Leif, and Simmons 2014). We use an improved version of the p-curve analysis that is more robust to errors and certain kinds of p-hacking (Simonsohn, Simmons and Leif 2015).

Moderators

Considering the characteristics of the studies, we selected five possible moderators that could influence the results since they are all related to methodological aspects or participants' characteristics: duration of the intervention, academic area of the study context, design of the study, grade of the participants and the outcomes under study.

Outcomes were divided into knowledge, which involves the measurement of the acquired knowledge level, and complex skills, which can never be fully automated and rely on/encompass knowledge available in long-term memory. Table 1 represents the moderators characteristics: name, description, type and values.

Table 1

Moderators characteristics

Name	*Description*	*Type*	*Values*
Duration	Duration of the intervention	Dichotomous variable	Short (less than one month) Long (more than one month)
Area	Academic area in which the study was conducted	Categorical variable	The studies were categorized based on Fan and Chen's (2001) characterization for academic areas: Math/quantitative Reading/language arts Science Social studies Other (e.g. music) Unspecified
Design	Design of the study	Categorical variable	Experimental Quasi-Experimental Non-Experimental
Grade	Grade of the participants	Dichotomous variable	Higher Education High School
Outcome	Dependent variable	Dichotomous variable	Knowledge Complex Skills

Results

Table 2 presents the synthesis of the 12 selected studies. Columns EG and CG represent the number of subjects in experimental and control groups, respectively. For the studies that had one single group, only the N column was fulfilled with the total number of participants. The Group column synthesizes the treatment applied into experimental groups versus control groups and the procedures to non-experimental groups.

Table 2

Characteristics of the selected studies

Study	*EG*	*CG*	*N*	*Age*	*Duration*	*Type*	*Area*	*Country*	*Group*	*Design*	*Grade*	*Outcome*	*d*
Flores (2011)	16	19	35	16	1 day	PhD thesis	Math	USA	EG: adaptative tutorial designed with 4C/ID CG: non-adaptative tutorial	Quasi-experimental	High School	Performance (Problem Solving skills)	*F* =1.16 *d* = 0.376
Kolcu, Öztürkçü and Kaki (2020)	-	-	26	25	21 days	Paper	Dentists	Turkey	Participants learned endodontics procedures in a distance education program based on 4C/ID	Non-experimental (single group with pre-test and post-test)	Higher education	Performance (psychomotor skills)	t-test = 5.97 d = 1.26
Lim and Park (2012)	12	10	22	30	1 semester	Conference paper	Computer Science	USA	EG: Whole task – (instruction was designed based on the 4C/ID model) CG: Part task	Quasi-experimental	Higher education	Performance (knowledge test: preparing a grade test using the data provided)	*t* (20) = 2.28) d = 0.976

Lim, Reiser and Olina (2009)	25	26	51	21	1 day (2 x 60 min)	Paper	Computer Science	USA	EG: Whole task – (instruction was designed based on the 4C/ID model) CG: Part task (traditional instructions)	Quasi-experimental	Higher education	Performance (complex skills)	EG: 30.9 (4.36) CG: 24.6 (6.67) *F* (1, 47) = 15.87 *d* = 1.114
Martínez-Mediano and Losada (2017)	20	20	40	18-20	1 semester	Paper	Engineering (Physics)	Bulgaria	EG: Used a platform with a system based on 4C/ID CG: Traditional classroom environment	Quasi-experimental	Higher education	Performance knowledge	EG: 4.59(0.96) CG: 3.58(1.1) *d* = 0.983
Melo and Miranda (2015)	78	51	129	14	2 weeks (90 min/week)	Paper	Physics	Portugal	EG: Took classes based on 4C/ID CG: Took classes based on expository teaching	Quasi-experimental	High School	Performance (transfer test)	EG: 11.2(1.9) CG: 8.8(2.6) *d* = 1.09

Melo and Miranda (2014)	81	50	131	14	2 weeks (90 min/week)	Conference Paper	Physics	Portugal	EG: Took classes based on 4C/ID CG: Took classes based on expository teaching	Quasi-experimental	High School	Performance (transfer test)	EG: 0,55 (0,87) CG: -0,42 (1,01) *d* = 1.48
Peng, Wang and Sampson (2017)	-	-	29	3-year college students	5 weeks	Conference Paper	Computer Science	China	Participants learned computer programming through a system based on 4C/ID	Non-experimental (single group with pre-test and post-test)	Higher education	Performance (knowledge test)	Pre-test: 46.34(17.29) Post-test: 53.31(15.38) *d* = 0.43
Rosenberg-Kima (2012)	31	33	64	21	1 day (105 min)	PhD thesis	Computer Science	USA	EG: Learned with task-centered instructions (4C/ID) CG: Learned with topic-centered instructions (traditional)	Non-experimental (single group with pre-test and post-test)	Higher education	Performance (skill-development)	EG: 8.96(2) CG: 7.51(2.87) $F(1,63) = 5.442$ *d* = 0.58

Sarfo and Elen (2007)	41	41	82	18	6 x 90 min	Paper	Engineering	Ghana	EG: 4C/ID learning environment without ICT CG: traditional method	Quasi-experimental	High School	Performance	EG: 8.84(3.12) CG: 5,44(3.46) *d* = 1.03
van Es and Jeuring (2017)	75	55	129	9-12	3-5 weeks (5 lessons)	Conference Paper	Computer Science (programming)	Netherlands	EG: Learned through 4C/ID instruction CG: Learned through constructionism instruction	Quasi-experimental	Elementary School	Performance (programming skills)	*t* (128) = 1.123 *d* = 0.019
Zhao, Wang and Yin (2017)	80	80	160	N/A	1 year	Conference Paper	Health (Nursing)	China	EG: Operating Room Nurse training with 4C/ID CG: Operating Room Nurse training with conventional training	Quasi-experimental	Higher education	Performance (knowledge test)	Chi-Square = 5.0 d = 0.34

From Table 2 we verify that all studies lasted between a day and a year. This leads us to consider that the effect of the 4C/ID model has in most cases been measured with short experiences and that longer ones may have a greater impact on student learning. However, it is not clear whether the previous studies' duration refers to the experience as a whole or just the time devoted to the task. Since this data is relevant for analysis, it would be important for studies to distinguish both types.

The studies were conducted in three distinct grades: elementary school (van Es and Jeuring 2017), high school (Flores 2011; Melo and Miranda 2015, 2014; Sarfo and Elen 2007) and higher education (Kolcu et al. 2020, Lim and Park 2012; Lim et al. 2009; Martínez-Mediano and Losada 2017; Peng et al. 2017; Rosenberg-Kima 2012). There were also some differences in the number of subjects in each study. Zhao et al. (2017) have the bigger sample. They conducted their study with 160 students. On the other hand, Lim and Park (2012) only had a sample of 22 students. Since the process of meta-analysis considers the number of participants in calculating the effect sizes, those differences were considered in the results.

From Table 2 we also verify that the studies were conducted in distinct domains. Mostly in Computer Science (Lim and Park 2012; Lim et al. 2009; Peng et al. 2017; Rosenberg-Kima 2012; van Es and Jeuring 2017), but also in Physics (Melo and Miranda 2015, 2014), Math (Flores 2011), Engineering (Martínez-Mediano and Losada 2017, Sarfo and Elen 2007) and Health (Kolcu et al. 2020, Zhao et al. 2017). In the next section (Sensitivity Analysis section), we study whether these differences influence the effect size.

More than half of the studies were conducted with young adults in a higher education environment. The remaining worked with youth and children. All studies used tests to quantify the performance of the subjects. However, the dependent variable, or outcome, is presented as skills acquisition (both complex and procedural) or knowledge. Since 4C/ID were mostly created for learning complex skills, in the Sensitivity Analysis section we will analyze whether these differences on the outcome under study will bias the effect size.

We also verify some differences in the design of each study which must be further analyzed by a sensitivity analysis (see next section). The studies were conducted under a quasi-experimental design (Flores 2011; Lim and Park 2012; Lim et al. 2009; Martínez-Mediano and Losada 2017; Melo and Miranda 2015, 2014; Sarfo and Elen 2007; van Es and Jeuring 2017; Zhao et al. 2017) or a non-experimental design (Kolcu et al. 2020; Peng et al. 2017; Rosenberg-Kima 2012).

In Table 3 we present the effect size of the 12 selected studies, their confidence interval (95%) and their weight in meta-analysis.

Table 3
Effect sizes of the 12 studies

	d	95%-CI	%W(random)
Flores (2011)	0.4100	[-0.2368; 1.0568]	6.9
Kolcu et al. (2020)	1.2600	[0.6720; 1.8480]	7.5
Lim and Park (2012)	0.9800	[0.1176; 1.8424]	5.2
Lim et al. (2009)	1.1100	[0.5220; 1.6980]	7.5
Martinez-Mediado and Losada (2017)	0.9900	[0.3432; 1.6368]	6.9
Melo and Miranda (2015)	1.0900	[0.7176; 1.4624]	9.6
Melo and Miranda (2014)	1.4800	[1.0880; 1.8720]	9.4
Peng et al. (2017)	0.4300	[-0.0796; 0.9396]	8.2
Rosenberg-Kima (2012)	0.5800	[0.2272; 0.9328]	9.8
Sarfo and Elen (2007)	1.0300	[0.5792; 1.4808]	8.8
van Es and Jeuring (2017)	0.0200	[-0.3132; 0.3532]	10.0
Zhao et al. (2017)	0.3400	[0.0264; 0.6536]	10.2

Table 3 and Figure 3 present the effect size of the selected studies. These results suggest a high effect of the use of 4C/ID model on knowledge acquisition (d = 0.79 [0.50 ;

1.08] which, according to Coe (2002) we can expect that between 76% and 79% of the subjects of the experimental group will perform better than the average of the control group.

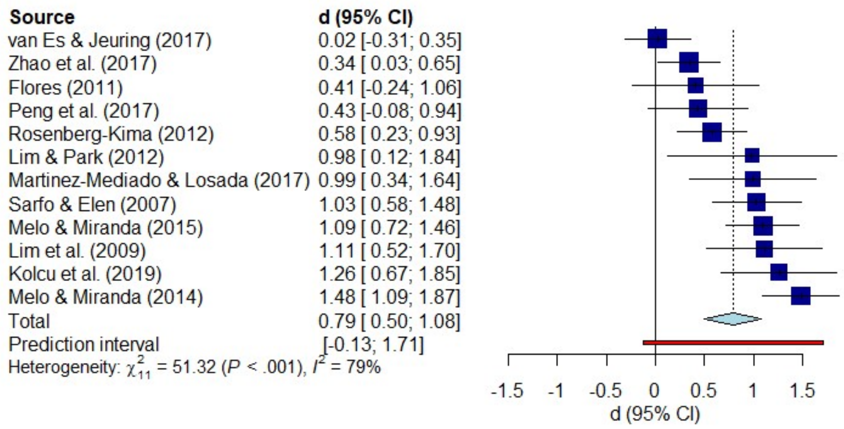

Figure 3. Forest plot with the selected studies.

However, the heterogeneity test revealed a high degree of heterogeneity (I^2 = 78.6% [63.1% ; 87.6%] ; Q = 51.32 ; p < 0.0001) which means that the moderators may have had some effect (Borenstein et al. 2009).

Sensitivity Analysis

We tested each moderator to check whether they are related to effect size differences (Borenstein et al. 2009). To achieve this goal, we checked the statistical significance of each moderator performing a t-test on β-weight of each moderator (Harrer, Cuijpers, Furukawa and Ebert 2019).

To ensure that our meta-regression was robust, we performed an intercorrelation matrix between the moderators to check for high correlations (Table 4).

Table 4

Intercorrelation matrix with the moderators.

	duration	area	design	grade	outcome
duration	1.00000000	0.1807016	0.09759001	0.6390644	0.2390457
area	0.1807016	1.00000000	0.00000000	0.2245444	0.4724556
design	0.09759001	0.00000000	1.00000000	-0.2425356	0.4082483
grade	0.63906444	0.2245444	-0.2425356	1.00000000	-0.2970443
outcome	0.23904572	0.4724556	0.4082483	-0.2970443	1.00000000

As we can see in Figure 4, Duration and Grade moderators are highly correlated. Since the correlation is significant and does not show the presence of a latent variable, we decided to exclude the duration of the intervention (Harrer et al. 2019) since in most of the 12 studies it is not possible to distinguish the time and frequency, which may lead to erroneous conclusions.

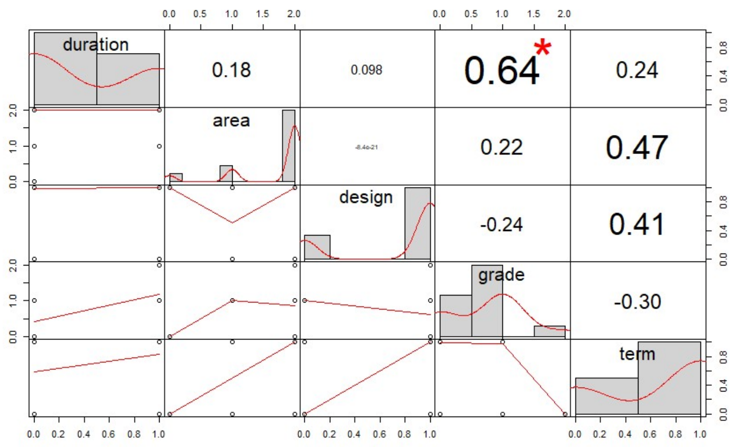

Figure 4. Intercorrelation between moderators

The moderator Area, in which studies were conducted, did not have a significant influence on the effect size ($p > 0.05$). The moderator Outcome was also not statistically significant ($p > 0.05$), suggesting that 4C/ID is suitable for both knowledge and skills (both complex and procedural) development. Also, the Design of the study did not demonstrate any evidence of significant influence on the effect size ($p > 0.05$).

On the other hand, the moderator Grade ($p < 0.05$) was statistically significant explaining, respectively, 76.22% of the heterogeneity. To assess the fit of our model, we performed a permutation test for the significant moderator. This test allows us to understand if we captured a real pattern in our analysis by re-calculating the p-values of the meta-regression (Harrer et al. 2019). The permutation test was significant ($p < 0.05$) suggesting that the grade of the participants indeed influences the effect size of the studies. The full results of the meta-regression and the permutation test are available in supportive information.

Discussion

Our study aimed to understand the effects of educational programs developed with the 4C/ID model on student performance.

Looking at the forest plot of Figure 3 and at Table 3, we can observe that there is a set of studies that reveal their effect sizes higher than the average value (Kolcu et al. 2020; Lim and Park 2012; Lim et al. 2009; Martinez-Mediano and Losada 2017; Melo and Miranda 2015, 2014; Sarfo and Elen 2007). We found that in this set of studies the effect of using educational programs developed with the 4C/ID model produced more pronounced effects on the participants' performance. How can we explain this difference?

From the sensitivity analysis, we conclude that the divergence in the studies were due to the academic area in which the study was conducted, the design and the outcome under study explained only a low percentage of the heterogeneity. This means that they have a nonsignificant influence on the effect size, indicating that the 4C/ID model might be helpful for developing educational programs that have positive effects on both knowledge and complex skills in a vast set of academic areas. The difference between knowledge and skills and their integration in a competent action, proposed by the 4C/ID model, has its origins in the history of psychology. The cognitive psychology of information processing established the difference between declarative knowledge and procedural knowledge (cf. Anderson 1983), the former associated with semantic memory and the latter with procedural memory. Hence it can be concluded that a model like 4C/ID, which aims to answer the problem of integrating knowledge and complex skills in a given domain, could not fail to consider these two dimensions. Furthermore, it is this integration that facilitates the transfer of learning.

We also verified that the selected studies present a correlation between their duration and the grade of subjects. However, since the studies were not detailed enough about the description of the duration of the intervention and the whole duration of the study, we could not be more comprehensive in the analysis. This limitation of our meta-analysis leads us to consider that studies in this area may be more accurate on the duration variable. In future work, an in-depth analysis on the effect of the duration of the intervention with 4C/ID should answer this issue.

Analysing each of the studies, we also found that in all, except Martínez-Mediano and Losada (2017) study, the researchers used validated instruments to evaluate the students' performances and, thus presented reliable measures. In the study reported by Martínez-Mediano and Losada (2017) there is no reference of procedures for determining the reliability of the knowledge assessment test. However, as several experimental and control groups were set up over time in two semesters, it can be inferred that the researchers may have been careful and done some analysis of the reliability of the questionnaire used in the performance evaluation.

The meta-analysis also evidences that the grade of the subjects has a significant influence on the effect size, having more powerful values in higher education than in lower grades. That is the reason that explains most of the heterogeneity in this selection. This was the case in the study reported by the van Es and Jeuring (2017), that had presented the lower effect size. This study was conducted in two different elementary schools and therefore with a different curricular organization, which may explain the magnitude of effect close to zero. These results are in line with van Merriënboer's work (2019) in which he argues that the 4C/ID model is more suitable for designing instruction for professional learning. But these results also lead us to raise two questions: are the lessons learned at these education levels less complex than those of certain professional activities, if we take into account the children's level of development? Or is it that the model is not suitable for these learnings and should be adapted? Learning to read, to write, to calculate and to think logically are complex learning and demand the integration of knowledge, attitudes and complex skills. They take time and a lot of training. Therefore, we are led to think that perhaps the lack of significant results of the aforementioned study are due to the non-adaptation of the model to these levels of education. But these questions will only be fully answered with further research.

Conclusion

The use of the 4C/ID model has been increasing in several areas, contributing to the enhancement of online (e.g. Frerejean et al. 2019; Sluijsmans, Prins and Martens 2006) and face-to-face learning environments (e.g. Costa 2019; Deep, Murthy and Bhat 2020). Despite not having met the inclusion criteria mostly due to not having published data on performance effects related to the use of the whole model, there has been high quality research that focuses on this model in several domains of activity such as health and medical education (e.g. Maggio, Cate, Irby, and O'Brien 2015; Tjiam, Schout, Hendrikx, Scherpbier, Witjes, and Van Merriënboer 2012), information-problem solving (e.g. Wopereis, Frerejean, and Brand-Gruwel 2015), application in higher educational programs (e.g. Frerejean et al. 2019) or even to study other possible related variables such as socio-economic status (e.g Costa and Miranda 2019) or socio-demographic factors (e.g. Posta and White 2017).

For future work, developing longer-term studies that use the complete model in learning environments in upper secondary, vocational and higher education students should be researched. This meta-analysis suggests that educational programs developed with the 4C/ID model help to develop educational programs that have a significant impact on student performance in several academic domains and it should be more widely used in universities for learning complex skills.

More robust methodologies will also be required, especially quasi-experimental designs with pre/post-test measures and experimental/control groups. The presence of measures that allow the calculation of the effect size will allow this meta-analysis to become more robust. Another important point is that researchers should be careful to use valid and reliable measuring instruments. On working these aspects, it will become possible to gather more detailed evidence on how specific design guidelines of the 4C/ID model contribute to the improvement of complex learning.

References

4CID.org (2020). Retrieved from https://www.4cid.org/home

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge: Harvard University Press

Anderson, J. R. (1993). *Rules of the mind*. NJ: Erlbaum.

Baddeley, A. (1997). *Human memory. Theory and practice* (revised ed.). UK: Psychology Press.

Blessing, S. B., & Anderson, J. R. (1996). How people learn to skip steps. *Journal Experimental Psychology: Learning, Memory, Cognition,* 22, 576–598.

Bransford, J., Brown, A., & Cocking, R. (2000). *How people learn: Brain, mind, experience and school* (expanded ed.). Washington D.: National Academy Press.

Bruner, J. S. (1966). *Toward a theory of instruction*. Cambridge MA: The Belnap Press of Harvard University Press.

Coe, R. (2002). It's the Effect Size, Stupid. What Effect Size Is and Why It Is Important. *Paper Presented at the British Educational Research Association Annual Conference*, Exeter, 12-14.

Conboy, J. (2003). Algumas medidas típicas univariadas da magnitude do efeito. *Análise Psicológica, 2*(21), 145–158.

Connelly, K., & Bruner, J. (Eds.) (1974). *The growth of competence.* USA: Academic Press.

Costa, J. (2019). Microworlds with different pedagogical approaches in introductory programming learning: effects in programming knowledge and logical reasoning, *Informatica, 43*, 145–174. doi: 10.31449/inf.v43i1.2657

Costa, J., & Miranda, G. L. (2019). Using Alice software with 4C/ID model: Effects in programming knowledge and logical reasoning. *Informatics in Education, 18*(1), 1–15. doi: 10.15388/infedu.2019.01

Deep, A., Murthy, S., & Bhat, J. (2020). Geneticus Investigatio: a technology-enhanced learning environment for scaffolding complex learning in genetics, Research and Practice in Technology Enhanced Learning, 15. doi: 10.1186/s41039-020-00145-5

Flores, R. (2011). *Examining the Design and Impact of Adaptively Faded Worked Examples on High School Students´ Mathematics Problem Solving Skills* (Doctoral Dissertation, Texas Tech University, Texas, USA). Retrieved from: https://ttu-ir.tdl.org/handle/2346/ETD-TTU-2011-08-1755

Frerejean, J., van Merriënboer, J. J. G., Kirschner, P. A., Roex, A., Aertgeerts, B., & Marcellis, M. (2019). Designing instruction for complex learning: 4C/ID in higher education. *European Journal of Education, 54*(4), 513-524

Harrer, M., Cuijpers, P., Furukawa, T.A, & Ebert, D. D. (2019). Doing Meta-Analysis in R: A Hands-on Guide. https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/. Accessed 01 June 2020. https://doi.org/10.1111/ejed.12363

Kester, L., Kirschner, P. & van Merriënboer, J. (2004). Information presentation and troubleshooting in electrical circuits. *International Journal of Science Education, 26*(6), 239-256. doi: 10.1080/69032000072809.

Kester, L., Kirschner, P. A., van Merriënboer, J. J. G., & Baumer, A. (2001). Just-in-time information presentation and the acquisition of complex cognitive skills. Computers in Human Behaviour, 17, 373-391.

Kokko, A. K., & Hirsto, L. (2021). From physical spaces to learning environments: processes in which physical spaces are transformed into learning environments. *Learning Environments, 24*, 71-85. https://doi.org/10.1007/s10984-020-09315-0

Kolcu, M. İ. B., Öztürkçü, Ö. S. K., & Kaki, G. D. (2020). Evaluation of a distance education course using the 4C-ID model for continuing endodontics education. *Journal of dental education, 84*(1), 62-71.

Lenhard, W. & Lenhard, A. (2016). *Calculation of Effect Sizes*. Retrieved from: https://www.psychometrica.de/effect_size.html. Dettelbach (Germany): Psychometrica. DOI: 10.13140/RG.2.2.17823.92329

Lim, J. & Park, S. (2012). *An Instructional Method for Competency-based e-Learning: A Whole-task Approach*. In P. Resta (Ed.), Proceedings of Society for Information Technology and Teacher Education International Conference 2012 (pp. 580-585). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE).

Lim, J., Reiser, R. & Olina, Z. (2009). The effects of part-task and whole-task instructional approaches on acquisition and transfer of a complex cognitive skill. *Learning Environments Research*, 57, 61-77. doi: 10.1007/s11423-007-9085-y.

Maggio, L. A., Cate, O. T., Irby, D. M., & O'Brien, B. C. (2015). Designing evidence-based medicine training to optimize the transfer of skills from the classroom to clinical practice: Applying the four component instructional design model. *Academic Medicine, 90*, 1457– 1461. https://doi.org/10.1097/ACM.0000000000000769

Martínez-Mediano, C., & Losada, N. R. (2017). Internet-Based Performance Support Systems in Engineering Education. *Revista Iberoamericana de Tecnologias del Aprendizaje*. 12(2), 86-93.doi: 10.1109/RITA.2017.2697778.

Mayer, R. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.). *The Cambridge Handbook of Multimedia Learning* (2nd Ed.) (pp. 43-71). Cambridge: Cambridge University Press.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. Journal of Educational Psychology, 86, 389-401.

Melo, M., & Miranda, G. L. (2014). Applying the 4C-ID model to the design of a digital educational resource for teaching electric circuits: Effects on student achievement. In H. M. Fardou, and J. A. Gallud (Eds). Proceedings of the 2014 Workshop on Interaction Design in Educational Environments (IDEE'14) (pp. 12-18). New York: ACM. http://doi.acm.org/10.1145/2643605

Melo, M., & Miranda, G. L. (2015). Learning electrical circuits: The effects of 4C/ID instructional approach in the acquisition and transfer of knowledge. *Journal of Information Technology Education:Research, 14*, 313-337. DOI: 10.28945/2281

Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development. 50*(3), 43-59.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D.G. (2009). *Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement*. Annals of Internal Medicine, 151, 264-269. 10.7326/0003-4819-151-4-200908180-00135

Nadolskyi, R., Kirschner, P. & van Merriënboer, J. (2004). Optimizing the number of steps in learning tasks for complex skills. *British Journal of Educational Psychology, 75*, 223-237. doi:10.1348/000709904X22403.

Nadolskyi, R., Kirschner, P. & van Merriënboer, J. (2006). Process support in learning tasks for acquiring complex skills in the domain of law. *Learning and Instruction, 16*, 266-278. doi:10.1016/j.learninstruc.2006.03.004.

Nisser, U. (2014/1967). *Cognitive Psychology*. New York: Psychology Press Classic Edition.

Paas, F. (1992). Training strategies for attaining transfer of problem solving skill in statistic: a cognitive load approach. *Journal of Educational Psychology, 84*(4), 429-434. doi: 10.1037/0022-0663.84.4.429.

Piaget, J. (1992). *Réussir et comprendre* (2ème ed.) Paris: PUF

Formatou: Inglês (Reino Unido)

Peng, J., Wang, M. & Sampson, D. (2017). Visualizing the complex process for deep learning with an authentic programming project. *Educational Technology* and Society, 20, 275-287.

Postma, T. C., & White, J. G. (2017). Socio-demographic and academic correlates of clinical reasoning in a dental school in South Africa. *European Journal of Dental Education, 21*(1), 58-65.

Reigeluth, C.M., Beatty, B.J., & Myers, R.D., (Eds.) (2017). *Instructional-Design Theories and Models, Volume IV: The learner-centered paradigm of education.* New York: Routledge

Reiser, R. (2001). A history of instructional design and technology. Part 1: A history of instructional media. *Educational Technology Research and Development, 49*(1), 53-64.

Rosenberg-Kima, R. (2012). *Effects of Task-Centered vs. Topic-Centered Instructional Strategy Approaches on Problem Solving – Learning to program in Flash*. (Doctoral Dissertation, Florida State University, Tallahassee, USA). Available from http://fsu.digital.flvc.org/islandora/object/fsu%3A183092.

Sarfo, F. & Elen, J. (2007). Developing technical expertise in secondary technical schools: The effect of 4C/ID learning environments. *Learning Environments Research, 10*, 207-221. doi: 10.1007/s10984-007-9031-2.

Skinner, B. F. (1954). The science of learning and the art of teaching. *Harvard Educational Review, 24*(2) 86-97.

Sluijsmans, D., Prins, F., & Martens, R. (2006). The design of competency-based performance assessment in e-learning. *Learning Environments Research, 9*, 45-66. doi: 10.1007/s10984-005-9003-3

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*(2), 261-292

Tjiam, I. M., Schout, B. M., Hendrikx, A. J., Scherpbier, A. J., Witjes, J. A., & van Merriënboer, J. J. (2012). Designing simulator-based training: an approach integrating cognitive task analysis and four-component instructional design. *Medical teacher, 34*(10), e698-e707.

van Es, N. & Jeuring, J. (2017). Designing and comparing two Scratch-based teaching approaches for students aged 10-12 years. Proceedings of Koli Calling 2017. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3141880.3141883.

van Merriënboer, J. J. (2019). *The Four-Component Instructional Design Model: An overview of its main design principles*. Maastricht, The Netherlands: Maastricht University.

van Merriëenboer, J. J., Clark, R. E., & De Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational Technology Research and Development*, 50(2), 39-64. https://doi.org/10.1007/BF02504993

van Merriënboer J., & de Croock, M. (1992). Strategies for computer-based programming instruction: Program completion versus program generation. *Journal of Educational Computing Research, 8,* 365– 394.

van Merriënboer, J., Jelsma, O., & Paas, F. (1992). Training for reflective expertise: A four-component instructional design model for complex cognitive skills. *Educational Technology Research and Development, 40*(2), 23-43. https://doi.org/10.1007/BF02297047

van Merriënboer, J., & Kirschner, P. (2001). Three worlds of instructional design: State of the art and future directions. *Instructional Science, 29*, 429-441.

Formatou: Português (Portugal)

Formatou: Português (Portugal)

van Merriënboer, J. J. G., & Paas, F. (2003). Powerful learning and the many faces of instructional design: Towards a framework for the design of powerful learning environments. In E. De Corte, L. Verschaffel, N. Entwistle, & J. J. G. van Merrie¨nboer (Eds.), *Powerful learning environments: Unravelling basic components and dimensions* (pp. 3–20). Oxford, UK: Elsevier Science.

Vygostky, L. (1978). *Mind in society. The development of higher order processes*. Cambridge, Mass.: Harvard University Press.

Wopereis, I., Frerejean, J., & Brand-Gruwel, S. (2015). Information problem solving instruction in higher education: A case study on instructional design. In S. Kurbanoglu, J. Boustany, S. Špiranec, E. Grassian, D. Mizrachi, and L. Roy (Eds.), *Information literacy: Moving toward Sustainability*. (Vol. 552, pp. 293– 302). Basel, Switzerland: Springer.