CRIStAL, Centre de Recherche en Informatique Signal et Automatique de Lille

Université de Lille – Faculté des Sciences et Technologies

Itec, Centre de Recherche Interdisciplinaire de l'Université de KU Leuven et de l' Imec

KU Leuven- Faculté des Sciences Psychologiques et Pédagogiques

Thèse en vue de l'obtention du titre de

DOCTEUR EN INFORMATIQUE et PEDAGOGIE

Learning Analytics pour la compréhension des processus d'apprentissage dans les environnements d'apprentissage en ligne

Charlotte Larmuseau

Soutenue le 30 Novembre 2020

<u>Rapporteurs:</u>
Prof. dr. Johan van Braak (Université de Gand)
Prof. dr. Pierre Beust (Université de Caen Normandie)
<u>Examinateurs:</u>
Prof. dr. Annelies Raes (KULeuven)
Prof dr. Jean Heutte (Université de Lille)
<u>Directeurs:</u>
Prof. dr. Fien Depaepe (KULeuven)
Prof. dr. Luigi Lancieri (Université de Lille)

-Let's keep it complex-

The current learning landscape is evolving in terms of what is learned and the context in which learning takes place. This can largely be related to the continuously changing requirements of today's labor market. More than ever, the importance is stressed of the use of rich authentic learning tasks that provide opportunities to acquire 21st -century skills such as complex learning. Complex learning is defined by van Merriënboer, Clark and De Croock (2002) as the integration and coordination of knowledge, skills, and attitudes that constitute real-life task performance, which enables the transfer of what was learned in school or training to daily life and work. Additionally, the availability of information technology has changed the traditional educational boundaries of time, space, and informational access. From an instructional design perspective, the combination of both phenomena poses a great challenge for researchers and instructional designers to implement instruction that meets the requirements of the current learning landscape (Ng, 2015). A research-based instructional design model that has proven to be effective in promoting complex learning is the four-component instructional design model (4C/ID-model; van Merriënboer et al., 2002). Nonetheless, offering an online learning environment based on a research-based instructional design model is not necessarily a guarantee for its effectiveness. As the learner is an active agent in the online learning process, the effectiveness of learning environments largely depends on student cognitive and motivational-affective characteristics.

In order to investigate characteristics that can influence the effectiveness of a 4C/ID-based online course and how effectiveness can be facilitated, the current research project was divided into respectively research track 1 and 2. On the basis of three studies, research track 1 examined the influence of students' cognitive and motivational-affective characteristics on use and learning outcomes. More particularly, Study 1 and 2 investigated the influence of students' technology acceptance and students' perceptions towards instructional quality. Additionally, Study 3 investigated the influence of students' prior knowledge and motivational characteristics. Findings of Study 1 and 2, reveal the importance of students' technology acceptance and perceived instructional quality on respectively the quantity and quality of use and students' learning outcomes. Moreover, mixed findings between Study 1 and 2 indicate results can be context-dependent (i.e. ecological validity). Additionally, findings of Study 3 indicate that students' prior knowledge and task value can influence differences in use. Furthermore, students' prior knowledge and differences in use positively influence students' learning outcomes. As a result, research track 1 indicates that individual differences influence the effectiveness of a 4C/ID-based online course.

Nonetheless, former research indicates that the influence of individual differences can be monitored by aligning the learning environment with students' learning needs (Moos & Azevedo, 2009). In order to align the online course with students' learning needs we should be able to detect learning process during online complex learning. Consequently, research track 2 explores in two studies whether physiological measures such as skin response measures (Study 4 and 5) and cardiovascular measures (Study 5) can be used to assess cognitive load during the online problem-solving process. Findings of Study 4 reveal that electrodermal activity (EDA) can be linked to self-reported cognitive load and that changes in cognitive load can be detected by EDA when differences in cognitive load are high. Findings of Study 5 appear to indicate that cognitive overload induces stress which was assessed via skin temperature (ST) and heart rate (HR). Both studies clarify that cognitive load is a complex concept to measure. In order to have insight into cognitive load, the studies of research track 2 emphasize the need for future studies to combine different measures in order to have more robust information on the mental effort exerted by a person during a cognitive task.

-Let's keep it complex-

L'éducation change en termes de ce qui est appris et de contexte dans lequel l'apprentissage a lieu. Cela peut être en grande partie lié aux exigences en constante évolution du marché du travail actuel. Plus que jamais, on souligne l'importance de l'utilisation de tâches d'apprentissage riches et authentiques qui offrent des possibilités d'acquérir des compétences du XXIe siècle telles que l'apprentissage complexe. L'apprentissage complexe est défini par van Merriënboer, Clark et De Croock (2002) comme l'intégration et la coordination des connaissances, des compétences et des attitudes qui constituent l'exécution de tâches réelles, ce qui permet le transfert de ce qui a été appris à l'école ou en formation dans la vie quotidienne et au travail. En outre, la disponibilité des technologies de l'information a modifié les limites traditionnelles de l'éducation en matière de temps, d'espace et d'accès à l'information. Du point de vue de la pédagogie, la combinaison de ces deux phénomènes représente un grand défi pour les chercheurs et les pédagogues qui doivent mettre en œuvre une pédagogie répondant aux exigences du contexte actuel de l'apprentissage (Ng, 2015). Le modèle de conception pédagogique à quatre composantes (modèle 4C/ID ; van Merriënboer et al., 2002) est un modèle de conception pédagogique basé sur la recherche qui s'est avérée efficace pour promouvoir l' apprentissage complexe. Néanmoins, offrir un environnement d'apprentissage en ligne basé sur un modèle de conception pédagogique basé sur la recherche n'est pas nécessairement une garantie de son efficacité. Comme l'apprenant est un agent actif dans le processus d'apprentissage en ligne, l'efficacité des environnements d'apprentissage dépend largement des caractéristiques cognitives et motivationnelles-affectives de l'étudiant.

Afin d'étudier les caractéristiques qui peuvent influencer l'efficacité d'un cours en ligne basé sur les 4C/ID et la manière dont l'efficacité peut être facilitée, le projet de recherche actuel a été divisé en deux pistes de recherche. Sur la base de trois études, la première piste de recherche a examiné l'influence des caractéristiques cognitives et motivationnelles-affectives des étudiants sur l'efficacité des environnements d'apprentissage en ligne. Plus particulièrement, les études 1 et 2 ont examiné l'influence de l'acceptation de la technologie par les étudiants et la perception de la qualité de l'enseignement par les étudiants. En outre, l'étude 3 a examiné l'influence des connaissances antérieures et des caractéristiques motivationnelles des élèves sur les différentes utilisations des composantes et les résultats de l'apprentissage. Les résultats des études 1 et 2 révèlent l'importance de l'acceptation des technologies par les étudiants et de la perception de la qualité de l'enseignement sur respectivement la quantité et la qualité de l'utilisation et les résultats d'apprentissage des étudiants. En outre, les résultats de l'étude 3 indiquent que (1) les connaissances antérieures et la motivation intrinsèque des étudiants peuvent influencer les différences d'utilisation et que (2) les connaissances antérieures des étudiants et les différences d'utilisation des composantes influencent positivement les résultats d'apprentissage des étudiants. Par conséquent, la piste de recherche 1 indique que les différences individuelles peuvent influencer l'efficacité d'un environnement d'apprentissage en ligne.

Néanmoins, des recherches antérieures indiquent que l'influence des différences individuelles peut être modérée en alignant l'environnement d'apprentissage sur les besoins d'apprentissage des étudiants (Moos & Azevedo, 2009). Afin d'aligner le cours en ligne sur les besoins d'apprentissage des étudiants, nous devrions être en mesure de détecter le processus d'apprentissage au cours de l'apprentissage complexe en ligne. Par conséquent, la deuxième piste de la recherche a examiné dans deux études si des données physiologiques liées à la peau (études 4 et 5) et les données physiologiques cardiovasculaires (étude 5) peuvent être utilisées pour évaluer la charge cognitive pendant le processus de résolution de problèmes en ligne. Les résultats de l'étude 4 révèlent que l'activité électrodermique (EDA) peut être liée à la charge cognitive auto-déclarée et que les changements de charge cognitive peuvent être détectés par l'EDA lorsque les différences de charge cognitive sont élevées. Les résultats de l'étude 5 semblent indiquer que la surcharge cognitive induit un stress qui a été évalué via la température de la peau (ST) et la fréquence cardiaque (HR). Les deux études précisent que la charge cognitive est un concept complexe à mesurer car elle change rapidement au cours du processus d'apprentissage et en raison de son interaction avec les états mentaux connexes.

-Let's keep it complex-

Het huidige onderwijslandschap evolueert voortdurend op vlak van wat, waar en wanneer er geleerd wordt. Ten eerste is er de voortdurend veranderende arbeidsmarkt, die de nadruk meer dan ooit op complexe, authentieke leertaken legt om 21ste-eeuwse vaardigheden aan te leren. Daarnaast verandert de beschikbaarheid van informatietechnologie de traditionele grenzen van tijd, ruimte en toegang tot informatie in het onderwijs. De combinatie van beide fenomenen vormt een uitdaging voor onderzoekers en onderwijskundige ontwerpers om een vorm van onderwijs te implementeren die aan alle eisen van het huidige onderwijslandschap voldoet (Ng, 2015).

Complex leren wordt door van Merriënboer, Clark en De Croock (2002) gedefinieerd als (1) de integratie en coördinatie van kennis, vaardigheden en attitudes en (2) de transfer van wat er geleerd werd naar het dagelijkse werk en leven. Dit complex leren wordt gestimuleerd door het 4C/ID-model, een onderwijskundig ontwerpmodel dat uit vier componenten bestaat (4C/ID-model; van Merriënboer et al., 2002). Toch is een op dit ontwerpmodel gebaseerde online leeromgeving niet noodzakelijk een garantie voor succes: omdat de lerende grotendeels verantwoordelijk is voor het online leerproces, is de effectiviteit van leeromgevingen afhankelijk van zijn cognitieve en motivationeel-affectieve karakteristieken. Daarom werd dit onderzoeksproject opgesplitst in twee onderzoekslijnen die nagaan welke kenmerken belangrijk zijn voor de effectiviteit van de online leeromgeving en hoe we die kenmerken kunnen bijsturen.

In onderzoekstraject 1 onderzochten we in drie studies hoe de cognitieve en motivationeel-affectieve karakteristieken van studenten het gebruik van een online leeromgeving voor complex leren en de leeruitkomsten van de studenten beïnvloeden. Studies 1 en 2 behandelden de invloed van enerzijds de attitude van studenten tegenover technologie en anderzijds de perceptie die ze hebben op de kwaliteit van het onderwijskundig ontwerp. Uit deze studies blijkt dat positieve attitudes en percepties een positieve invloed hebben op zowel kwantiteit als kwaliteit van het gebruik van de online leeromgeving en (daaruit volgend) de leerresultaten. Studie 3 onderzocht de invloed van voorkennis en motivatiekenmerken van studenten op het verschillende gebruik van een online leeromgeving en op de leeruitkomsten. We zagen dat voorkennis en intrinsieke motivatie invloed kunnen hebben op verschillen in gebruik en dat de voorkennis en het verschillend gebruik van de componenten (in functie van de leernoden) de leerresultaten positief beïnvloeden. De resultaten van de eerste onderzoekslijn tonen dus aan dat individuelle verschillen een invloed hebben op de effectiviteit van een 4C/ID-gebaseerde online leeromgeving.

Voorgaand onderzoek bevestigde al dat individuele verschillen kunnen worden bijgestuurd door de leeromgeving op de leerbehoeften van studenten af te stemmen (Moos & Azevedo, 2009), maar hiervoor moeten we wel in staat zijn de leerprocessen te meten. In onze tweede onderzoekslijn onderzochten we in twee studies of fysiologische data, zoals huidgeleiding- en temperatuur (Studies 4 en 5) en cardiovasculaire reacties (Studie 5), gebruikt kunnen worden om de mentale belasting tijdens het online probleemoplossingsproces te meten. Resultaten uit Studie 4 tonen aan dat huidgeleiding (EDA) inderdaad aan zelf-gerapporteerde mentale belasting kan gekoppeld worden en dat EDA grote veranderingen in mentale belasting detecteert. Studie 5 wijst erop dat mentale overbelasting kan gemeten worden aan de hand van huidtemperatuur (ST) en hartslagfrequentie (HR). Beide studies tonen de bijzondere complexiteit aan van het meten van de mentale belasting tijdens het leerproces.

"If everyone is moving forward together, then success takes care of itself." - Henry Ford

Dit proefschrift is het resultaat van vier jaar intensief werken. Maar nog belangrijker, het is ook het resultaat van een goede entourage. Daarom wil ik eerst en al vooral iedereen bedanken die een bijdrage heeft geleverd aan de totstandkoming van dit proefschrift.

Fien Depaepe*.* Bedankt dat je mijn promotor wilde worden en zijn. Je hebt mij (her)opgeleid tot een echte kritische onderzoeker. Je gaf mij vanaf het begin al veel vertrouwen. Dit zorgde ervoor dat ik veel vrijheid kreeg om mij bij te scholen en te verdiepen in onderwerpen die ik interessant vond. Je gaf me de ruimte om te kunnen vallen en opstaan. Daarnaast stond je altijd klaar voor een overleg of nodige bijsturing. Je snelle en *constructieve* feedback was zeer waardevol. Ik ga in ieder geval wat moeten *afkicken* van het idee dat ik mijn werk bij jou kan aftoetsen. Zoals Marie voorheen al had bevestigd, ben je inderdaad een fantastische promotor.
Het was ook fijn dat we af en toe eens konden praten over de dagdagelijkse dingen. Ik denk dat we qua karakter wel een match zijn, wat heel bevorderend was voor onze samenwerking.

Piet Desmet*.* Je bent *de bruggenbouwer* van mijn onderzoeksproject. Je hebt mij een unieke kans aangeboden om als pionier een *cotutelle* te starten aan de Université de Lille. Deze samenwerking was een grote meerwaarde voor mijn proefschrift aangezien mijn onderzoek zich bevindt op het snijpunt van de pedagogische wetenschappen en computerwetenschappen.
Ik wil hier aansluitend ook Frederik Cornillie, Ine Windey en Antoine Besnehard (I-SITE; Université Lille Nord-Europe) bedanken om mij hierin te sturen en mijn *cotutelle* (KU Leuven- Université de Lille) op te volgen.

Luigi Lancieri**.** Je vous remercie beaucoup de m'avoir laissé participer à votre groupe de recherche (CRIStAL). J'étais très fière de pouvoir m'inscrire dans votre université. Cela m'a permis d'expliquer mes recherches à vos collaborateurs et de participer à des séminaires très intéressants dans votre faculté. De ce fait, j'ai pu me former une meilleure idée de la perspective de l'informaticien. Je suis également très reconnaissante de votre accueil toujours aussi chaleureux. Nous, les Flamands, avons encore des choses à apprendre à ce sujet. Je suis convaincue qu'il aurait été intéressant de poursuivre les recherches à base des données vidéo, mais malheureusement, le Covid-19 a changé notre rythme de travail.

Dank aan **alle leden van de begeleidingscommissie** : Jan Elen, Wim Van Den Noortgate, Katrien Verbert en Johan van Braak (Universiteit Gent). Door met jullie in discussie te gaan en jullie feedback te verwerken, is de kwaliteit van mijn proefschrift aanzienlijk verbeterd.

Dank aan **alle juryleden:** Annelies Raes, Johan van Braak (Ugent), Pierre Beust (Université de Caen Normandie) en Jean Heutte (Université de Lille). Hartelijk dank om mijn proefschrift zo grondig te lezen

en voor het (ongetwijfeld) stellen van interessante en kritische vragen tijdens mijn verdediging. *Je voudrais également remercier tous les membres du jury d'avoir lu mon doctorat en profondeur et d'avoir posé des questions intéressantes et critiques lors de ma soutenance.*

Alle partners *die mijn onderzoek mogelijk hebben gemaakt*.
Dank aan Hendrik Coucke (VIVES, Kortrijk), Griet Grymonpon (VIVES, Brugge), Lucie Martin en Marieke Pieters (*RHIZO* Lyceum OLV Vlaanderen, Kortrijk) om mee te werken aan mijn onderzoek.
Dank aan *Jan Cornelis* (Imec) om met mij samen te werken rond de data van de Imec-wearables.
Dank ook aan de mensen van MICT (Jessica Morton, Klaas Bombeke, Bram Van Acker)- ik heb veel bijgeleerd over onderzoek met fysiologische data door deze samenwerking.
Daarnaast wil ik zeker ook *Jan Elen* bedanken. Niet alleen voor de feedback die ik kreeg op mijn eerste en derde artikel, maar ook voor de vele inspirerende artikels waarop een groot deel van mijn theoretisch kader is gebaseerd.
Bedankt ook aan *Jeroen van Merriënboer*. Enerzijds om mij te ontvangen aan de Maastricht University; anderzijds omdat het 4C/ID-model de basis is van het eerste deel van mijn onderzoekstraject.

Mijn (ex-)collega's. En dat zijn er best veel – dank aan alle collega's uit Lille (CRIStAL), Kortrijk (Itec) en Leuven (CIP&T).
In deze tijden is het duidelijk dat collega's belangrijker zijn dan ooit. Ik wil hier toch een paar mensen in de kijker zetten.
Marie, dankjewel om mij te *triggeren* om te doctoreren. Ondanks de soms wat onduidelijke communicatie ben ik zeker dat je mijn West-Vlaamse tongval (en 'humor') mist. *Annelies*, dankjewel om zoveel mijn (hypersociale) conference-buddy te willen zijn- leuke manier om de mensen van Universiteit Gent te leren kennen! Dank aan *Louise*, voor de vele gesprekken op ons bureau en mijn *Moodle*-buddy te zijn; *Pieter*, om mijn *R*-buddy te zijn; *Stefanie*, om de 10 miles te lopen met me; *Jo en Tine*, voor alle steun en tips; *Marlou, Esther,* om mij te entertainen tijdens de studiereis: *Juliet, Bert, Kim, ...* voor de grappige lunchpauzes; *Sylke, Bieke en Ele*, voor de koffiepauzes, enz.

Vrienden en vriendinnen. Ik wil ook kort mijn vrienden bedanken. Die hebben ervoor gezorgd dat ik niet altijd bezig was met mijn onderzoek. En geloof me, dat was soms meer dan welkom!
Ook de mensen van *Sailing Venues*, want die zeilreizen waren voor mij steeds fantastische ervaringen - '*Life is better on a boat!'*

Familie. Ik wil uiteraard *mijn ouders* bedanken. Mijn ouders hebben mijn studies mogelijk gemaakt en gesteund in al mijn keuzes. Ik heb nu eenmaal de liefste ouders. Ik kan er altijd terecht voor een babbel. Mijn ouders zijn ook heel geïnteresseerd in wat ik doe. Mijn mama zou zelfs naar de andere kant van de wereld reizen om één van mijn presentaties bij te wonen (LAK19 - Sydney, Australië).
Daarnaast wil ik ook mijn *broer en schoonzus* bedanken. Ik heb (en nog steeds) altijd enorm naar hen opgekeken. Mijn broer is nu eenmaal *mijn grootste held*. Ik was bijvoorbeeld enorm trots toen ik

mocht werken bij Kulak, aangezien mijn broer daar had gestudeerd- dat kon niet anders dan *de beste universiteit ooit* zijn.
De verdediging van mijn schoonzus was dan weer *de eerste verdediging* die ik ooit in mijn leven had bijgewoond. Ik ben blij dat ik nu in haar voetsporen kan treden.
Bovendien ben ik ook trotse meter van hun prachtige kinderen: *Clemens, Catharina & Julius*.

Mijn gezin. *Julia*, jouw lach doet me altijd smelten. Ik ben elke dag onder de indruk van wat jij al kan. Ik ben zo trots op jou- en dat zal altijd zo blijven! *Diederik,* mijn kapitein, jouw ambitie werkt inspirerend en heeft ook mij gemotiveerd om net dat tikkeltje vérder te gaan- onder jouw motto *'Failure is not an option'*

Charlotte, november 202

Chapter One

Introduction: theoretical background of the research project

1. Theoretical background of the research project

Nowadays, learners have to be prepared to deal with continuously evolving environments and rapidly changing demands which shape the fast-developing world of the 21st-century (Ng, 2015). Consequently, more attention is given to the provision of courses that have real-world relevance and utility and provide opportunities for acquiring 21st-century skills such as complex learning (Jonassen, 2000; Lane & D'Mello, 2019). Moreover, due to the ubiquitous presence of IT, an increasing amount of courses in higher education take place online (Ng, 2015). From an instructional design perspective, the combination of both phenomena poses a great challenge for researchers and instructional designers on how to design online courses that promote complex learning. Consequently, this research project aimed at defining factors that can improve the effectiveness of online courses for complex learning. In this introduction, we will elaborate on the theoretical framework of the research project. Additionally, an overview of the studies and their research aims is provided.

1.1. Complex learning

This section firstly defines complex learning. Secondly, it elaborates on students' characteristics that can influence the effectiveness of complex learning. Thirdly, it provides information on how learning environments can be designed to promote complex learning.

1.1.1. Defining complex learning

The current learning landscape is changing. Nowadays, learning activities should provide opportunities for acquiring 21st-century skills such as complex learning (Jonassen, 2000; Lane & D'Mello, 2019). Complex learning is defined by van Merriënboer et al. (2002) as (1) the integration and coordination of qualitatively different knowledge, skills, and attitudes that constitute real-life task performance and (2) transferring what is learned in daily life and work. The definition of complex learning formulated by van Merriënboer et al. (2002) can be related to the definition of complex problem solving defined by Sweller (1994). Sweller (1994) claims that the complexity of learning tasks or problems can be defined by *element interactivity*. That is, the number of elements and their interrelationships that simultaneously need to be processed. Sweller (1994) indicates that an element is anything that needs to be learned or processed such as a definition, a formula, a sub-goal of the problem, etc. The more novel elements a problem or learning task contain, the higher the load on working memory will be (Jonassen, 1997;

Sweller, 2010). Both definitions indicate that the complexity of what has to be learned is highly determined by the number of (novel) elements that need to be processed simultaneously (van Merriënboer & Sweller, 2005).

1.1.2. Challenges for complex learning related to students' characteristics

The effectiveness of complex learning is strongly related to students' cognitive and motivational-affective characteristics. This section will discuss how these characteristics can influence the effectiveness of complex learning.

Cognitive characteristics

Solving complex problems can be quite challenging for the learners' working memory as it requires the learner to coordinate and integrate multiple elements simultaneously (Jonassen, 2000; Sweller, 2010; van Merriënboer, 2013). Moreover, the absence of prior domain knowledge, puts an even heavier burden on the learner's working memory during complex problem solving (Schwaighofer, Bühner & Fischer, 2017). This interplay between students' prior knowledge and working memory can be explained by *the architecture of human memory* (Sweller, 2010). The architecture of human memory makes a distinction between *long-term* and *short-term memory* or *working memory* (van Merriënboer & Sweller, 2005). The working memory has two prominent characteristics that are critical for instructional design. A first characteristic of the working memory is the limited capacity when dealing with new information. More particularly, the working memory can only process approximately seven pieces of novel information at a time. When working memory capacity is exceeded (i.e. cognitive overload), additional information will be lost. A second characteristic of the working memory is the limited duration of information being processed. More particularly, information in working memory can only be stored for no more than a few seconds. Moreover, all information is lost after about 20 seconds, unless it is refreshed by rehearsal (Kirschner, Kester & Corbalan, 2011; Sweller, van Merriënboer & Paas, 2019). As a result, working memory capacity plays an important role in the ability to deal with high element interactivity. Nonetheless, working memory has fewer limitations when dealing with domain-specific prior knowledge retrieved from *long-term memory*. Domain-specific prior knowledge is defined by knowledge structures present in long-term memory, also known as *cognitive schemas*. These schemas are connected elements, which are processed as single elements in the working memory. As such, the presence of cognitive schemas reduces the number of elements that have to be held and processed in the working memory. Subsequently, the presence of domain-specific cognitive schemas deduces cognitive load during complex learning (Sweller et al., 2019). Given this interplay between domain-specific prior knowledge and the learners' working memory capacity, optimizing cognitive load is considered to be important when designing learning environments for complex learning (Kalyuga & Singh, 2015).

Motivational- affective characteristics

The importance of motivational-affective variables for the effectiveness of the learning process can be among others explained by the *Expectancy-Value Theory.* The Expectancy-Value Theory emphasizes the importance of academic self-efficacy and task value. *Academic self-efficacy* is defined as the individual's perception of his or her current capacity to execute behaviors necessary for accomplishing a task (Zimmerman, 2000). When dealing with complex tasks, efficacious learners tend to persist, cope, and adapt well, even when their prior domain knowledge is low. In contrast, learners with low confidence in their ability can become frustrated, overwhelmed, and demotivated (Taipjutorus, Hansen, & Brown, 2012). *Task value* is more related to the learners' intrinsic motivation. More particularly, task value captures reasons for engagement by identifying how much the learner values the desired outcome (Duncan & McKeachie, 2005). It is believed that learners will learn more from complex tasks when the information is relevant for them (Liem, Lau & Nie, 2008).

Furthermore, former studies have indicated that *the perceived instructional quality* of a learning environment influences learners' academic engagement and subsequently course performance. More particularly, Frick, Chadha, Watson, and Zlatkovska (2010) conducted a study in which they investigated the association between students' perceived instructional quality based on the five *First Principles of Instruction* of Merrill (2002) and students' course achievement. Merrill's (2002) five First Principles for Instruction define that learning is promoted when: (a) learners are engaged in solving real-world problems, (b) existing knowledge is activated as a foundation for new knowledge, (c) new knowledge is demonstrated to the learner, (d) new knowledge is applied by the learner and, (e) new knowledge is integrated into the learner's world. Several empirical studies have indicated that students who agreed that the First Principles of Instruction were incorporated in the learning environment, were more likely to achieve high levels of mastery of course objectives (Frick et al., 2010; Martens, Bastiaens & Kirschner, 2007).

1.1.3. Designing learning environments for complex learning

Many current instructional design models indicate that complex learning can be promoted when learning activities are problem-centered and lead students through *four distinct phases of learning* (a) activation of prior knowledge, (b) demonstration of knowledge, (c) application of knowledge, and (d) integration of knowledge into the learners' world (Frick et al., 2010; Jonassen, 1997; Merrill, 2002). On that basis, Merrill (2002) developed the five *First Principles of Instruction,* that were mentioned in section 1.1.2. An instructional design model that incorporates these first principles of instruction and has proven to be effective for designing learning environments for complex learning is *the 4C/ID model* developed by van Merriënboer et al. (2002). The basic claim of the 4C/ID model is that all environments for complex learning can be described in terms of four interrelated components *(1) learning tasks, (2) part-task practice, (3) supportive information, and (4) procedural information*. *Learning tasks* are meaningful complex tasks, based on real-life situations and require the integration of both recurrent (i.e. routine) and non-recurrent (i.e. novel, effortful) subskills. Non-recurrent skills are more complex and involve schema-based problem solving, reasoning, and decision making, whereas recurrent skills are domain-specific rules, step-by-step procedures, etc. *Supportive information* helps learners to perform the problem-solving and reasoning aspects of the learning tasks (i.e. non-recurrent subskills). *Procedural information* is step-by-step instruction and helps students to perform routine aspects of learning tasks (i.e. recurrent subskills; van Merriënboer & Sluijsmans, 2009). *Part-task practice* provides additional practice for selected recurrent subskills in order to reach a required level of automaticity (e.g. exercises on applying a grammar rule).

Former studies investigated the effectiveness of the 4C/ID-model, by comparing a 4C/ID-based learning environment with more conventional learning methods. Lim, Reiser, and Olina (2009) compared the effects of whole-task training and part-task training on the acquisition and transfer of a complex cognitive skill for novices and advanced learners. They found that both novices and advanced learners achieved better at whole-task performance when they received whole-task training. Additionally, the study of Sarfo and Elen (2005) compared three groups who had to learn how to design a single building plan based on local conditions. The control group was taught according to a traditional approach that was applied in technical schools, whereas the experimental groups were taught according to the 4C/ID-approach. Results revealed that the experimental groups outperformed the control group on the post-test. Similarly, the study of Melo and Miranda (2014) aimed at exploring the effectiveness of a 4C/ID-based learning environment in terms of reproduction and transfer of learning. The results of the experiment supported the hypothesis that a 4C/ID-based learning environment contributed more positively to learning of concepts (related to theme “Electric Circuits”) compared with a learning environment based on a conventional (objectives-based) instructional method.

The effectiveness of the 4C/ID-model for complex learning is the result of *whole-task training* (van Merriënboer & Kirschner, 2018). In contrast, conventional learning methods are known for an objectives-based approach that breaks down the (authentic) learning tasks into their constituent parts. Subsequently, instructors choose an instructional method for each of those separate parts and their corresponding objectives. As a result of this teaching method, students do not learn to coordinate these separate parts into a coherent whole which leads to fragmentation of the learning material. Consequently, students only have limited opportunities to train and execute the whole-task in an integrated manner (e.g. when doing an internship; Frerejean et al., 2019a).

However, a disadvantage of a 4C/ID-based learning environment is that rich authentic whole-tasks are high in element interactivity (Jonassen, 2000, Sweller, 2010). As a result, a risk of the whole-task approach is that a high amount of cognitive load can be induced, which might hamper learning (van Merriënboer & Sluijsmans, 2009). This phenomenon can be explained within the framework of The Cognitive Load Theory (CLT). CLT provides insight into cognitive load, by distinguishing intrinsic, extraneous, and germane load (Sweller, 2010). *Intrinsic load* is determined by the element interactivity of a particular task. *Extraneous load* is cognitive load that has nothing to do with achieving the learning objectives. More particularly, extraneous load can be induced by suboptimal or complex instructional procedures. Moreover, it can be reduced by instructional guidance as this can ensure that the learner does not focus on irrelevant information (Sweller et al., 2010). *Germane load* is a productive cognitive load that contributes to learning and is closely related to intrinsic load (Chen & Kalyuga, 2020; Sweller et al., 2019; van Merriënboer & Kirschner, 2018). Given that intrinsic and germane load are largely related, more researchers plea for a dual-model of cognitive load that includes only intrinsic and extraneous load (Sweller et al., 2019). It is important to keep intrinsic and extraneous load within the capacity of the working memory. When the total amount of intrinsic and extraneous load exceeds the working memory, cognitive overload occurs (Chen & Kaluyga, 2020). In addition, underload means that not all available working memory is used. Conditions of both underload and overload should be prevented as these can obstruct learning (Sweller, 2010).

The 4C/ID-model aims at preventing cognitive underload and overload by the provision of four components as illustrated in *Figure 1* (van Merriënboer & Kirschner, 2018). The *intrinsic load* is controlled by starting each task class with a worked example, and additionally, by organizing the learning tasks in simple-to-complex task classes (Sweller et al., 2019). Learning tasks within a simpler task class contain fewer elements and fewer element interactivity and are therefore less complex. Gradually, as learners gain expertise, task classes increase in complexity (Sweller, 2010; van Merriënboer & Kirschner, 2018). By organizing the task classes in this manner the model prevents the possibility of working memory resource depletion (Chen & Kalyuga, 2020). *Both extraneous and intrinsic load* is managed by the provision of procedural, supportive information and part-task practice. These components can support learners in devoting fewer memory processes to irrelevant elements. Moreover, when cognitive schemes are constructed and automated, this frees up processing resources that the learner can devote to learning (Lee, Donkers, Jarodzka & van Merriënboer, 2019; van Merriënboer & Kirschner, 2018).

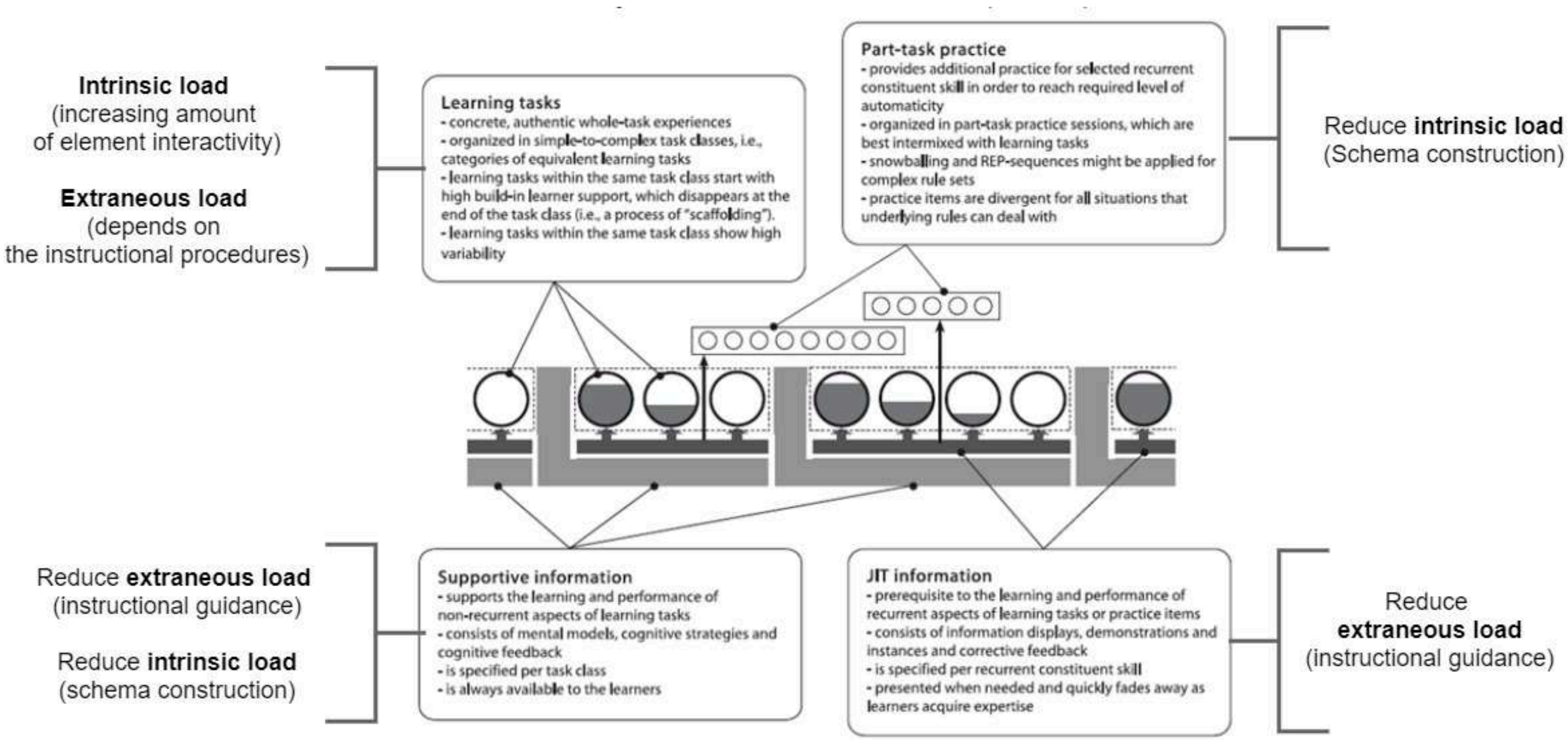


Figure 1.1.: Graphical overview of the four components (van Merriënboer et al., 2002, p. 44) in relation with CLT

1.2. Online learning environments for complex learning

1.2.1. Digitalization in educational settings

Digital technology is ubiquitous in modern society, and has changed the landscape of education (Lin, Parsons & Cockerham, 2019). Digital technologies, also referred to as *educational technologies* when applied for education purposes, include computers, laptops, tablets, smartphones, cameras, communication and collaborative resources, simulations, etc. (Jameson, 2013). Moreover, the availability of a huge range of Internet-dependent technologies such as open-source Learning Management Systems (LMSs), social networking sites and apps, cloud storage, educational games and resources sites, etc. have changed the traditional educational boundaries of time, space, and access to information.

Educational technology provides many benefits in educational contexts. These technologies (1) allow for *interactivity* as the technology can systematically respond to the actions of the learners (e.g. serious games), (2) can present learning material *adapted* to students' characteristics retrieved from trace data (e.g. Intelligent Tutoring Systems, ITS), (3) can provide *feedback* based on the quality of the learner's performance, (4) provide opportunities for learners to *self-regulate* their learning, (5) allow for *non-linear access*, that is, learners can select and receive learning activities that deviate from a set order, (6) can provide quick *connections between representations* of subject matter and, (7) allow for communication with a wide audience (e.g. peers, subject-matter experts, pedagogical agents, etc.; National Academy of Sciences, 2018).

Consequently, the omnipresence of educational technologies also creates the need to develop novel conceptual frameworks for effective pedagogy and strategies to enhance active, engaged, and meaningful learning that will lead to better learning outcomes for nearly every imaginable subject matter typical of 21st-century tasks (Ng, 2015).

1.2.2. Shaping online education

Online learning is one of the fastest-growing trends in the use of educational technology (Ng, 2015). Other terms used for online learning are e-learning, computer-based instruction, distance learning, technology-enhanced learning, etc. These terms are still expanding because of the evolving nature of applications of online learning (Bell & Federman, 2013). Regardless of the small differences between these terms, they can all be defined as learning within an educational learning environment in which the Internet is used as the infrastructure of Information Technology (IT; Lane & D'Mello, 2019; Ng, 2015). As such, online learning shares many characteristics with the Internet such as openness, accessibility, and interactivity (Bell & Federman, 2013).

Due to the increasing amount of technology tools, online learning can highly resemble face-to-face teaching. In online courses, students can watch videos, presentations, listen to audio files, and read or complete assignments at times of their convenience. Additionally, online learning allows for synchronous and asynchronous interactions and collaboration between students and the instructor via chat rooms, forums, and collaboration tools (e.g., Wikispaces, Skype, Google Drive, etc.; Ng, 2015). The majority of online learning in higher education takes place via LMS and/or virtual learning environments (VLEs)) of which Moodle and Blackboard, are the most well-known providers (Long & Siemens, 2011). An increasing amount of institutions in higher education are also offering Massive Open Online Courses (MOOCs). The most well-known providers of MOOCs are Coursera and edX. MOOCs are basically *online courses* consisting of short video lectures, online tests, and discussion forums. *Massive* refers to the capacity to enroll a large number of students and *Open* refers to free education as the learning materials are accessible for all users (Ng, 2015). As a consequence, learners have the opportunity to participate in lectures from famous professors of top universities. Nonetheless, despite these promising advantages, MOOCs are often criticized for offering a low-quality instructional design which frequently results in low completion rates (Phan, McNeil & Robin, 2016).

1.2.3. A 4C/ID-based online learning environment

As aforementioned in section 1.3. the 4C/ID model is a research-based instructional design model that has proven to be effective in promoting complex learning (Lim et al., 2009; Melo & Mirando, 2014; Sarfo & Elen, 2005). Former studies have also indicated that this model can easily be applied in online contexts (Frerejean et al., 2019a; Melo & Miranda, 2014). Nonetheless, to be effective, online learning requires learners to take responsibility for grasping learning opportunities (Elen, 2020).

This responsibility can partly be monitored by adjusting the amount of provided learner control in an online learning environment. *Learner control* is the extent to which learners have opportunities to make their own decisions regarding pace, strategies, or sequence according to their interests and preferences.

Consequently, learner control can be promoted in online courses when the learners are offered a set of learning tools and services which they can select and use in a way they deem fit (Väljataga & Laanpere, 2010).

Embedding learner control in online courses is strongly advised as it is believed to enhance learning. More particularly, former research indicated that giving learners *a sense of control* positively influences students' motivation, level of confidence, self-directed learning, and persistence to deal with the learning content (e.g. Dollinger, 2000; Lee, Choi & Kim, 2012; Taipjutorus, 2012). This phenomenon can also be related to Mayer's (2003) *pacing principle* indicating that learners learn better when they are allowed to have control over the pace of their learning.

Applied to the 4C/ID-model, learner control is largely promoted when the components of the instructional model are offered in a non-embedded manner and subsequently can be consulted voluntarily and non-linearly (van Merriënboer & Sluijsmans, 2009). For instance, students can consult procedural information or select part-task practices to remediate weaknesses in their achievements. Nonetheless, making thoughtful decisions concerning learning needs can be quite challenging and often depends on students' motivation and perceptions on the perceived functionality (see also: 1.2.2.; Elen, 2020; Lust, Juarez Collazo, Elen & Clarebout, 2012).

1.2.4. Technology acceptance of an online learning environment

Taking learning opportunities can be influenced by the learner's acceptance of the online learning environment as a useful learning tool (Schepers & Wetzels, 2007; Šumak, Hericko and Pušnik, 2011; Terras & Ramsay, 2015). A model that is frequently used to explain why an individual accepts or rejects IT, is *the Technology Acceptance Model* (TAM) proposed by Davis (1989). TAM postulates that the combination of *perceived usefulness* (PU) and *perceived ease of use* (PEOU) indicates students' technology acceptance. *PU* is defined as the degree to which the learner believes that using the IT will improve his or her learning performance, whereas *PEOU* is the degree to which the learner believes that using the IT will be user-friendly (Venkatesh & Davis, 2000). Former studies that used TAM as a baseline (i.e. by adopting PU and PEOU) have confirmed that technology acceptance influences the actual use in the context of online learning (Islam, 2013; Juarez Collazo, Wu, Elen & Clarebout, 2013; Šumak et al., 2011). Additionally, a former study that has extended TAM with students' learning outcomes, indicated that the actual use of the online course can also influence students' learning outcomes (Juarez Collazo et al., 2013). This seems to illustrate the importance of the perceived technological acceptance for the effectiveness of online learning environments.

1.2.5. The use of Learning Analytics to investigate (online) learning processes

The growing use of online learning environments in educational institutions has provided a wide range of trace data. More particularly, learning can be assessed at a fine level of granularity based on learners' actions, choices, and performance (e.g. clickstream data, time spent, time-on-task, time delays between responses, sequential movement etc.). The collection and analysis of trace data provide opportunities to obtain insight into the online learning processes and/or align the online learning environment with the learners' needs (Davies et al. 2017; Wong et al. 2019).

Two research fields, namely *educational data mining (EDM) and Learning Analytics (LA)* have a joint interest in how trace data can be exploited to benefit education and the science of learning (Baker & Inventado, 2014). EDM is mainly concerned with the analysis of large-scale educational data (i.e. also referred to as *Big Data*), e.g., retrieved from online learning environments. EDM uses complex computational methods (e.g. Machine and Deep Learning, Social Network Analysis) to model learning processes. These models can be used to predict student engagement, performance, drop-out rates, etc. Moreover, the models can be used to create sophisticated interactive learning systems (e.g. ITS, recommender systems) that deliver customized instruction and allow for adaption of the learning system according to the learner profile (i.e. student model; Lane & D'Mello, 2019; Wise & Schaffer, 2015). Consequently, EDM mainly focuses on the technical specifications of predictive approaches (i.e. algorithms and student models).

Nonetheless, when modeling learning processes, researchers should be careful when treating trace data as direct measures of learning (Davies et al., 2017). Although it may seem promising to predict learning behavior by using huge datasets, the sole focus on automated discovery without consideration of human judgment can be harmful (Davies et al., 2017; Gašević, Dawson & Siemens, 2015; Wise & Shaffer, 2015; Wong et al., 2019). From this demand, LA has emerged as a research field (Larusson & White, 2014; Selwyn, 2019). The official definition of LA is "the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environment in which it occurs" (Long & Siemens, 2011, p. 34). Accordingly, research in the field of LA aims at understanding and optimizing the learning process by providing guidelines on how to adapt educational opportunities in accordance with the learners' needs and abilities (Gašević et al., 2015).

Researchers from the Learning Analytics and Knowledge community (LAK) emphasize that a *research-based learning theory* should be taken into account during all stages of the study, to have insight into how trace data is related to learning processes, that is, when formulating hypotheses, when selecting important variables from trace data and, when interpreting the findings etc. (Davies et al., 2017; Milligan, 2018; Gašević et al., 2015; Wise & Schaffer, 2005; Wong et al., 2019). Additionally, Gašević et al. (2015) indicate that learners are *active agents* in their learning process. Consequently, learners that receive the same instructional conditions can choose different learning trajectories (Elen, 2020; Lust, Elen & Clarebout, 2013). To better understand the choices that learners make within online courses, LA-researchers should take into account students' *internal conditions* (e.g. prior knowledge) and *external conditions* (e.g. task complexity, online context) when investigating trace data (Gašević et al., 2015).

Nonetheless, information on internal and external conditions is often hard to capture solely via log-data (Moissa, Bonnin & Boyer, 2019; Sharma & Giannakos, 2020). Consequently, different sources of data can provide information on learning processes that are the result from internal and external conditions and their interrelationships (e.g. cognitive load). These different sources of data can include self-reported data or data retrieved from students' communication patterns (e.g. speech, writing, gaze), behavioral measures (e.g. blink frequency, head movements), physiological data (e.g. skin-response measures, cardiovascular measures) and neural patterns (e.g. electrical activity of the brain) etc. (Chen et al., 2016; Moissa et al., 2019). Insights extracted from multimodal data enables researchers to investigate learners' behavior in ways that would not be possible when using individual data sources. Moreover, former studies indicated that the combination of multimodal data leads to significantly better prediction models of learning outcomes (e.g. Smets et al., 2018b; Worsley, 2018). When researchers use multimodal resources to understand, predict, and optimize learning, these studies can be situated within the research field of Multimodal Learning Analytics (MMLA; Cukurova, Kent & Luckin, 2019).

1.3. Research aims and overview of the conducted studies

The hypothesized conceptual framework of the current PhD project indicates that the effectiveness of 4C/ID-based online courses largely depends on students' cognitive, motivational-affective characteristics and the use of a 4C/ID-based online course (Liem et al., 2008; Song, Singleton, Hill & Koh, 2004; Taub, Azevedo, Bouchet & Khosravifar, 2014). Moreover, it emphasizes the importance of the integration of a learning theory, and the connection of internal and external conditions with trace data to understand complex learning processes (e.g. Gašević et al., 2015; Wong et al., 2019).

1.3.1. Research track 1

Against this theoretical framework, a first research track investigated the influence of individual differences and use on the effectiveness of a 4C/ID-based online course. As aforementioned in section 1.3., a 4C/ID-based learning environment promotes complex learning (e.g. Sarfo & Elen, 2005). Nonetheless, the provision of a qualitative research-based learning environment by no means ensures that students will grasp learning opportunities (Elen, 2020). Given that learners are active agents of their learning processes, grasping learning opportunities strongly depends on their cognitive, and motivational-affective characteristics (Elen, 2020; Gašević et al., 2015). In the introduction, we provided an overview of characteristics that can influence the use of the online learning environment and subsequently, students' learning outcomes (i.e. prior knowledge, motivation, perceived instructional quality, technology acceptance). However, we do recognize that other characteristics (e.g. test anxiety), which were not covered within the research project, might influence the effectiveness of online learning environments for complex learning.

1.3.2. Studies of research track 1

The general aim of research track 1 was to investigate the influence of individual differences and the use on the effectiveness of an online learning environment for complex learning. Therefore, three studies were conducted in which two online courses were developed in Moodle, and systematically designed according to the research-based 4C/ID-model. In those studies we investigated relationships between students' cognitive and motivational-affective characteristics on the one hand and students' use of the learning environment and their learning outcomes on the other hand, as illustrated in *Figure 2.* The content of the online learning environments were 'teaching French as a foreign language'(i.e. Study 1) and 'learning French as a foreign language' (i.e. Study 2/3).

In **the first and second study,** we focused on students' *technology acceptance* and *the perceived instructional quality*. Study 1 examined how technology acceptance influences the quantity of use of the

4C/ID-based online learning environment and how this influences learning gain in a highly ecologically valid context. Study 2 investigated how the perceived instructional quality of a 4C/ID-based online learning environment influenced technology acceptance, as well as the quantity and the quality of use (i.e. course performance) of the online learning environment.

The third study was conducted in the same context of study 2 but had a different research focus. More particularly, this study focused on investigating the influence of students' *cognitive* and *motivational* characteristics on *differences in use*. Moreover, study 3 investigated the influence of the differences in use, controlled for students' cognitive and motivational characteristics, on students' learning outcomes.

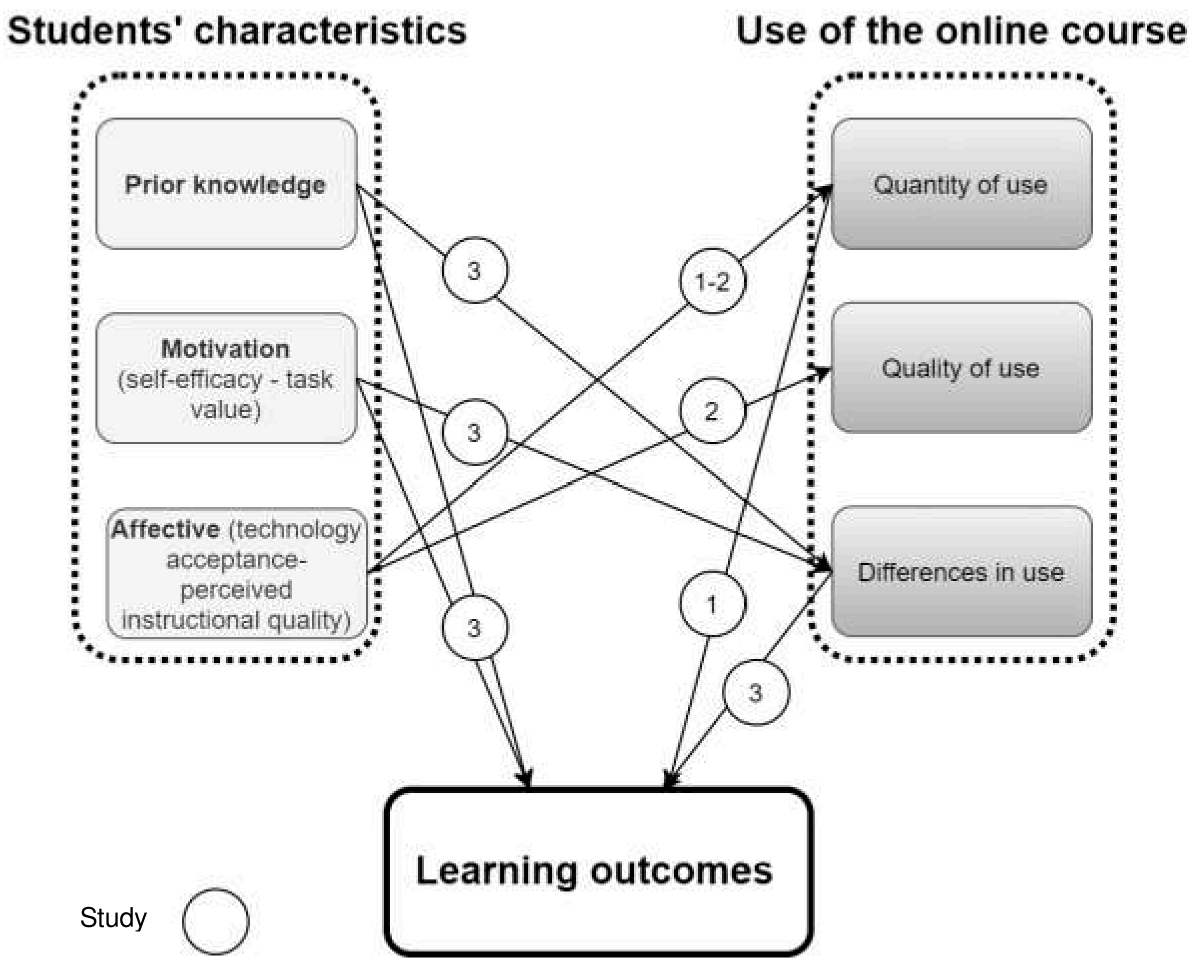


Figure 3.1: Overview of the studies of Research Track 1

By conducting these three studies in a comparable and systematized way, research track 1 strived to provide insight into the influence of individual differences on the effectiveness of a 4C/ID-based online course. Former studies have indicated that it is possible to monitor the influence of individual differences on the effectiveness of online learning environments by adjusting the learning environment to students' individual differences (e.g. Moos & Azevedo, 2009). As the current research project focuses on complex learning, it is advisable to provide customized support in line with students' cognitive functioning. Consequently, cognitive processes that are undesirable (e.g. cognitive underload or overload) can be prevented. In order to provide customized support, we should first and foremost be able to measure cognitive processes during online complex learning.

1.3.3. Research track 2

In order to be able to assess cognitive load during online complex problem solving, a second research track focused on investigating whether cognitive load can be measured via physiological data. The physiological approach is based on the assumption that changes in cognitive load can influence body properties (Moissa et al., 2019). The physiological measures that were explored were *skin responses* such as electrodermal activity (EDA) and skin temperature (ST) and *cardiovascular responses* such as heart rate (HR) and heart rate variability (HRV). This selection of physiological data was based on prior research on cognitive load assessment, the wearability and non-obtrusiveness of the technologies (i.e. wristband and chest patch).

1.3.4. Studies of research track 2

In research track 2, two online courses were developed in Moodle containing different learning tasks on geometry and statistics. In order to induce differences in cognitive load, task complexity of the learning tasks was manipulated based on element interactivity (i.e. intrinsic load) and the provision or absence of instructional guidance (i.e. extraneous load; Sweller, 2010). The design of the studies is illustrated in *Figure 3*.

In a fourth study differences in cognitive load were investigated during the problem-solving process of a high and low complex task on teaching geometry. In view of verifying the manipulation and investigating how changes in cognitive load can be measured, self-reported cognitive load (incl. intrinsic, extraneous, and germane) was combined with *EDA* and *ST* during the problem-solving process. Additionally, differences between self-reported and physiological data during the problem-solving process of the high and low complex tasks were examined. Moreover, in order to investigate how self-reported data is related to physiological data, this study correlated the self-reported cognitive load with physiological data. Finally, through visual analysis, peaks of EDA and ST during the online complex problem-solving process were associated with specific learning activities.

In **a fifth study** four sets of learning tasks on probability calculations in statistics were developed in an online learning environment. The four sets of learning tasks were manipulated based on two dimensions, to induce intrinsic load and extraneous load. In order to measure cognitive load during online problem solving, the self-reported cognitive load was combined with physiological data such as *EDA, ST, HR and HRV.* Firstly, the study examined differences across the four interventions in view of physiological data. Additionally, in order to investigate how self-reported data is related to physiological data, this study investigated how much variance of the self-reported cognitive load was explained by physiological data. Finally, in view of creating a more adaptive online learning environment based on students' cognitive load, a machine learning model was created in order to detect high cognitive load during online problem-solving.

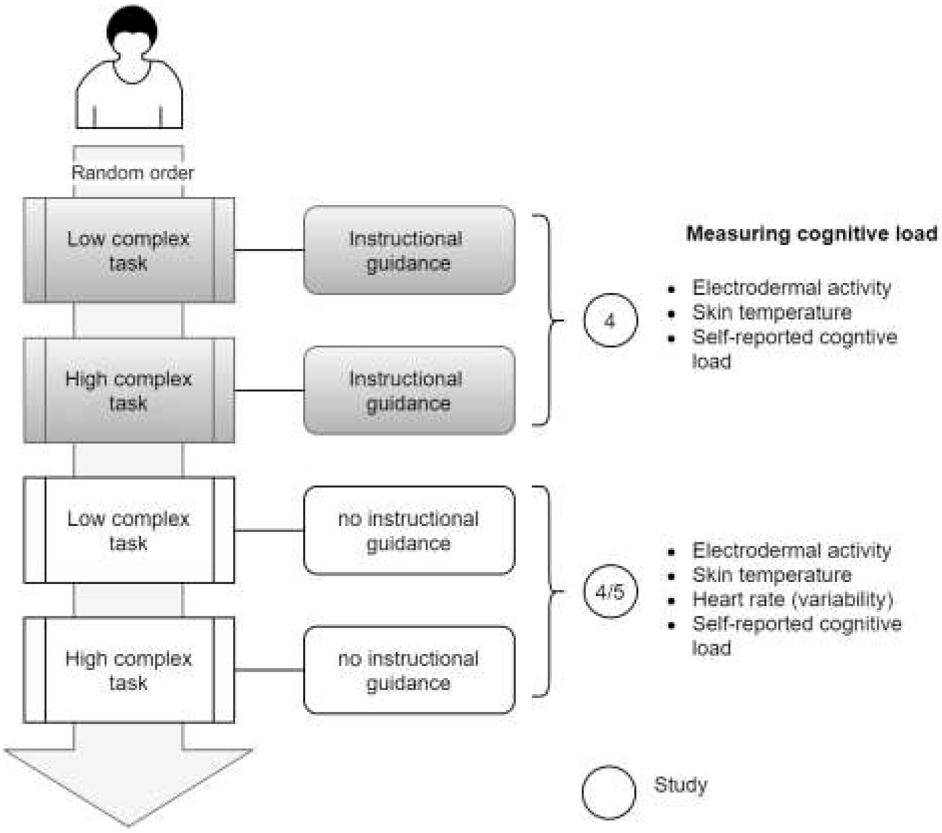


Figure 2.2: Overview of the studies of Research Track 2

By conducting these two studies in a comparable and systematized way, research track 2 strived to provide insight into whether (non-obtrusive) physiological data can be used for assessing cognitive load. The findings can be useful for follow-up research that focuses on predicting cognitive load and/or creating interactive online learning environments based on students' cognitive needs during online complex learning.

1.3.5. Overview of the doctoral thesis

The doctoral thesis is not written as a monograph but is a compilation of five studies that have been published. Therefore, some overlap in theoretical background and methodology exists among the various studies. Table 1 gives an overview of the structure of the doctoral thesis.

Table 1.1.

Visualization of the structure of the doctoral thesis

Investigating the effectiveness of online courses for complex learning	
Introduction	
Research track 1: Individual differences determining the effectiveness of a 4C/ID-based online course	
Study 1:	*Technology acceptance of a 4C/ID-based online course*
Study 2:	*The perceived instructional quality of a 4C/ID-based online course*
Study 3:	*The influence of cognitive and motivational characteristics on differences in use*
Research track 2: MMLA for measuring cognitive load during online complex learning	
Study 4:	*Combining physiological data and subjective measurements*
Study 5:	*Physiological data: a promising avenue for cognitive (over)load detection*
Discussion and concluding remarks	

A general conclusion and discussion of the five studies complete the doctoral thesis. In this section, we will discuss the findings per research track. Additionally, this will be completed by methodological and practical implications, combined with the limitations and perspectives for future research.

Chapter Two:

Technology acceptance

The text of this chapter was published as: Larmuseau, C., Evens, M., Elen, J., Van Den Noortgate, W., Desmet, P., & Depaepe, F. (2018). The relationship between acceptance, actual use of a virtual learning environment and performance: an ecological valid approach. *Journal of Computers in Education, 5,* 95-111. https://doi.org/10.1007/s40692-018-0098-9

2. Technology acceptance of a 4C/ID- based online course

2.1. Introduction

Even though the number of online courses being delivered in higher education is increasing, their effectiveness is often limited (Bell & Federman, 2013; Johnson & Aragon, 2003; Terras & Ramsay, 2015). A common problem seems to be that online courses tend to build on a very traditional view of learning (Johnson & Aragon, 2003; Revere & Kovach, 2011). This traditional view of learning sees the student as a passive receiver of knowledge and along with this view many online courses provide academic, decontextualized information and exercises that can be used primarily to drill and practice (Herrington, Oliver & Reeves, 2003). Meanwhile, the effectiveness of online courses can be greatly enhanced when it is designed properly by taking the educational objectives and the learners' needs into account (Elen, 2004; Huang, Rauch & Liaw, 2010). The instructional design should enable learners to transfer more complex cognitive skills or competencies (van Merriënboer et al., 2002). Moreover, instructional designers should be able to gain and maintain students' attention by providing a virtual learning environment (VLE) that stimulates engagement and participation. In order to achieve that goal, it is clear that the importance of understanding students' psychological constructs of motivation, metacognition and cognitive abilities that influence students' interactions with online courses should be recognized (Terras & Ramsay, 2015). As we recognize that students' characteristics and pedagogical considerations are interrelated, instructional designers need to employ approaches that increase student engagement in the learning environment and that allows us to achieve our educational objectives (Czerkawski & Leyman, 2016).

Merrill (2002) claims that student engagement can be promoted by incorporating the First Principles of Instruction, more particularly by (1) using real-world problems, (2) activating existing knowledge, (3) demonstrating new knowledge, (4) allowing students to apply the new knowledge and, (5) allowing the learner to integrate the new knowledge into the learners' world. An instructional design model that uses the First Principle of Instruction as the driving force for designing learning environments and that has proven to be especially useful for complex learning is the four-component instructional design model (4C/ID model) elaborated by van Merriënboer et al. (2002). The 4C/ID-model focuses on the whole task learning by providing four interrelated components, namely, learning task, part-task practice, supportive information, and just-in-time information. The design of the VLE studied is based on the 4C/ID-model.

The VLE aims at developing pre-service teachers' professional knowledge for teaching oral interaction French. Teaching oral interaction French is highly complex as it requires an instructor to combine domain-specific declarative and procedural knowledge with teaching skills and teacher attitudes. To deal with the complexity the learning tasks consist of authentic class situations, the part-task practice consists of drill-and-practice exercises (e.g. grammar and vocabulary), the supportive information is the bridge between what students already know and new knowledge (i.e. theory) and finally, just-in-time information is information that is prerequisite to the learning of the recurrent aspects of the learning tasks. Teacher professional knowledge for teaching oral interaction French can be defined as complex since it comprises the integration of knowledge, skills, and attitudes (Merrill, 2002; Kirschner & van Merriënboer, 2018).

Nevertheless, having access to technology, based on a well-considered instructional design aimed at improving student engagement, by no means ensures it will be used or used effectively (Thompson, Higgins & Howell, 1991). Students do not always embrace these learning opportunities, and therefore enhanced learning outcomes cannot be ensured (Clarebout, Horz, Schnotz & Elen, 2010). The successful implementation of the VLE highly depends on the students' perception of its usefulness and ease of use (i.e. useful in achieving their educational and personal goals and facilitating the learning process). These factors have been proven to affect students' initial acceptance and future usage of a VLE (Lau & Woods, 2009; Šumak et al. 2011; Tarhini, Hone & Xiaohui, 2013). In this study, the constructs perceived ease of use (PEOU) and perceived usefulness (PU) from the Technology Acceptance Model (TAM) were retrieved to examine the effect of students' acceptance on and the actual use, on the one hand, and the relationship between actual use of the VLE and their performance on the other hand. The following section clarifies how the 4C/ID-model is appropriate for complex learning and how TAM can contribute to understanding the impact of the students' acceptance of the VLE on the actual use, and the actual use in turn on performance.

2.2. Theoretical background

Instructional design for complex learning

The primary factor in any instructional initiative, regardless of format or venue, is the quality of the instructional design that is ultimately implemented (Johnson & Aragon, 2003). Many approaches, models, and frameworks exist for designing quality learning environments. The main concern is how to design a learning environment that conduces a high level of student engagement (Czerkwasky & Leyman, 2016). Based on a review of instructional design theories, Merrill (2002) identified five First Principles of Instruction to promote student engagement. He claims that student engagement will be promoted when (1) instruction is problem-centered (i.e. use of real-world problems), (2) relevant previous experience is activated, (3) students are exposed to demonstrations of what they are to learn, (4) students have opportunities to try out what they have learned with instructor feedback and, (5) learners are encouraged to integrate what they have learned into their personal lives. If one or more of these principles are violated during instruction, Merrill argues that learning can be negatively influenced (Frick et al., 2010).

One unified model that aligns with the above principles is the four-component instructional design model (4C/ID-model) elaborated by van Merriënboer et al. (2002). The 4C/ID-model involves four interrelated blueprint components: learning tasks, part-task practice, supportive and just-in-time information (van Merriënboer, 1997). The learning tasks are the backbone of the design process, and are authentic, whole-task experiences that confront the student with all constituent skills that make up a complex skill. They allow the simultaneous practice of domain-specific knowledge and cognitive strategies. They are clustered into task classes according to the level of their complexity or difficulty. The first task in each task class has high learner guidance and support in comparison to the others. The part-task practice consists of items that are provided to learners to improve recurrent aspects of the whole complex skill (i.e. learning tasks). The instructional methods of the part-task practice aim at reaching a very high level of automaticity. For each learning task supportive information is given. Supportive information provides all the information needed to complete non-routine aspects of learning tasks, which frequently involves problem solving and reasoning. It helps learners to relate what they already know with what may be helpful to know in order to solve the learning tasks. Just-in-time information is a prerequisite to the learning and performance of recurrent aspects of the learning tasks. It allows students to complete and learn routine aspects of learning tasks by specifying exactly how to solve the routine aspects of the tasks. It is presented just in time when learners need it (Kirschner & van Merriënboer, 2018). As aforementioned, the content of the VLE in the present study relates to teaching pre-service teachers how to prepare and conduct a lesson on oral interaction in French. Teaching oral interaction in French in the context of a foreign language is highly complex. While teaching, a teacher needs to use his or her pedagogical content knowledge (i.e. PCK), content knowledge (i.e. CK), and pedagogical knowledge (i.e. PK) simultaneously in various situations (Jonassen, 1997; van Gog, Sluijsmans, Joosten-ten Brinke & Prins, 2008). The 4C/ID-model intends to facilitate complex learning (van Merriënboer, Kirschner & Kester, 2003), therefore the instructional design of the VLE was based on the 4C/ID-model. However, even though a VLE is based on a qualitative instructional design, performance gain can be obstructed by students' unwillingness to accept and use VLE to improve their learning (Selim, 2003; Šumak et al., 2011). Therefore, in the next section, factors that influence students' acceptance of a VLE are discussed.

Technology Acceptance

Research on technology acceptance is very compendious, building on the basis of social psychology and sociology theories (Davis, 1989; Legris, Ingham & Collerette, 2003; King & He, 2006; Šumak et al., 2011; van Raaj & Schepers, 2008). Fishbein and Ajzen (1975) developed the Theory of Reasoned Action (TRA) that has proven to be successful in predicting and explaining the psychological determinants of behavior. According to TRA, the immediate determinant of a person's behavior is the intention to perform the behavior (BI) which is influenced by the attitude towards the behavior and subjective norm, which is a person's perception that he/she should perform that particular behavior (Venkatesh & Davis, 2000). Davis (1989) introduced an adaption of TRA, the Technology Acceptance Model (TAM). As TRA was designed to explain any human behavior, TAM was designed to explain computer usage behavior. The goal of TAM is to explain technology acceptance across a broad range of information technologies, while at the same time being both parsimonious and theoretically justified (Davis, Bagozzi & Warshaw, 1989). TAM has been empirically tested in many context and fields and has proven to be a useful theoretical model in helping to understand and explain the acceptance, adoption and use of information technologies (Legris et al., 2003). Moreover, Šumak, et al. (2011) systematically examined existing knowledge in the field of online learning acceptance and concluded that TAM is the most common ground theory in online learning acceptance literature. Davis (1989) claims that two cognitive constructs PU and PEOU are of primary relevance for technology acceptance (Davis et al., 1989).

It is hypothesized that PU and PEOU are major influences of an individual's attitude towards using technology (ATT), thus, ultimately relating to actual use (King & He, 2006; Schepers & Wetzels, 2009). PU is defined as the extent to which a person believes that using the system will enhance his or her performance in terms of the effectiveness of learning, productivity (i.e. time-saving). PEOU is defined as the extent to which a person believes that using the system would be free of effort, in terms of physical and mental effort (i.e. ease of learning; Teo & Zhou, 2016). ATT is hypothesized to be a major determinant of BI (Davis, 1989). However, prior research found that the mediating role of attitude toward using between PU and PEOU on BI has not always been confirmed (Burton-Jones & Hubona, 2006; van Raaj & Schepers, 2008; Venkatesh & Davis, 2000). For instance, Teo (2009) compared two versions of TAM: with attitude and without attitude and found that attitude towards use does not contribute to the total variance in usage. This study supported previous studies that found the attitude construct in TAM to be unnecessary. Furthermore, TAM is usually validated by using a measure of BI rather than actual use (Turner, Kitchenham, Brereton, Charters & Budgen, 2010). Therefore, Turner et al. (2010) investigated whether TAM explains actual use. A meta-analysis of 79 relevant empirical studies showed that BI is likely to be correlated with actual use, but PU is less likely correlated with actual use. Additionally, the associations were lower for objectively measured technology use (e.g. computer-recorded form) than for self-reported usage. Self-reported usage was in each study determined by asking two to four questions related to the frequency of use and/or the amount of time spent using the system. Former researchers have pointed out that self-reported usage is a simplistic view of system use that has important shortcomings. Self-reported usage is known to be subject to the common method bias, which distorts and exaggerates the causal relationship between independent and dependent variables (Benbasat & Barki, 2007; Legris et al., 2003; Turner et al., 2010).

Few prior researchers examined the direct effect of students' acceptance (i.e. PU and PEOU) on actual use in the context of VLEs, and those who did, used self-reported usage as the dependent variable (Islam, 2013; van Raaj & Schepers, 2008, Selim, 2003). Selim (2003) concluded that PU and PEOU (i.e. indirect effect) are key determinants for self-reported usage. The relationship between PU, PEOU and actual use was well supported, accounting for 83% of the total variance. Van Raaj and Schepers (2008), in accordance with Selim, found a direct effect of PU and an indirect effect of PEOU on self-reported usage (33% of the total variance). Islam (2013) studied the direct relationship of both constructs (i.e. PU and PEOU) on use, and found that both PU and PEOU had a direct effect on self-reported usage (45% of the total variance). Furthermore, using BI or self-reported use as the main dependent variable in prior research blinded researchers to other important effects, such as, students' performance (Benbasat & Barki, 2007). Juarez Collazo et al. (2014) were pioneers in studying the relationship between students' acceptance and the quantity of tool use in a computer-based learning environment, on performance. Tool use was defined as non-embedded support devices whose use depends on the students' initiative (i.e. using the tools depends on students' actions; Clarebout & Elen, 2006). In contrast with prior studies, Juarez Collazo et al. (2014) measured the quantity of tool use objectively (i.e. log files). This study indicates that PU has a positive influence on actual use (R^2 = .17) and actual use positively influences students' performance (R^2 = .11). This study indicates that when actual use is measured objectively, less variance is explained by the constructs PU and PEOU. Additionally, it points out the importance of students' PU and actual use on performance. However, an important limitation of the study of Juarez Collazo et al. (2014) was that the design of the study was merely experimental. The intervention was very short since the VLE consisted of only 10 different screens. Furthermore, the students possibly felt forced to use the tools during the intervention since they had to use them in a computer room under supervision. Additionally, the content of the computer-based learning environment was not integrated in their official educational program. In other words, the ecological validity of the study is rather low. To conclude, the majority of TAM research on VLE used self-reported actual use and had a technology focus. More specifically, it has been limited to experimental studies of adoption which differs from the real world in critical ways. Little research has placed the technology in its actual learning context (Benbasat & Barki, 2007; McGill & Klobas, 2009). Therefore, in the present study, the research design aims at maintaining the integrity of the real-life situation (i.e. ecological validity) in the experimental context and objectively measures actual use.

Research model and research questions

The present study aimed at investigating the relationship between students' performance and their acceptance and actual use of the VLE in the context of foreign language teacher education. Firstly, the relationship between students' acceptance to use the VLE and actual use was investigated. Taking TAM as a baseline, we used PU and PEOU as indicators of students' acceptance, and we assumed a direct effect of PEOU on PU. Based on previous research, we assumed that PU and PEOU affect actual use (Islam, 2013; Juarez Collazo et al., 2014; Selim, 2003; van Raaj & Schepers, 2008). Secondly, the effect of the students' acceptance and actual use of the VLE on students' performance was studied. Based on the proposed research model, i.e. *Figure 1*, we formulated the following research questions:

- *RQ 1*: How does students' acceptance (i.e. PU and PEOU) influence actual use of the VLE
- *RQ 2:* How does the actual use of the VLE influence students' performance?

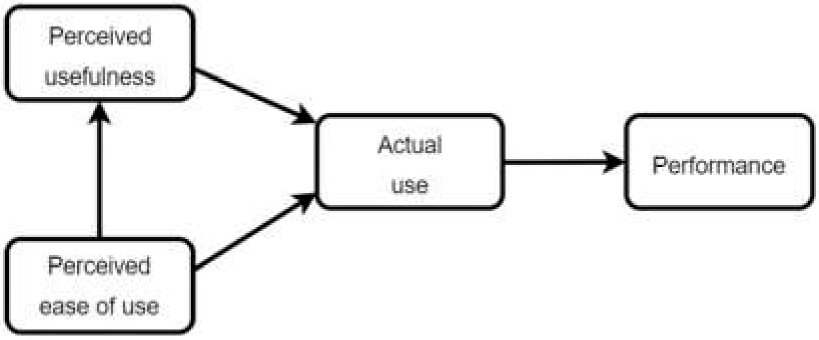


Figure 2.1.: The research model

2.3. Method

Participants

The study took place in the Flemish part of Belgium. In Flemish primary teacher training institutions, students are prepared to teach all courses that are part of primary education. French is one of the courses that is taught in primary education because it is one of the country's official languages. More specifically, French is compulsory starting from the fifth grade. Hence, French is part of the curricula in Flemish primary teacher training institutions. The content of the course was related to learning to teach oral interaction, in which aspects of PCK, CK and, PK were integrated. The participants were pre-service primary school teachers, recruited from two teacher training institutions i.e. second-year students. All of them had already experience in teaching French. The participants (N = 193) had an average age of 20 years. The majority of the participants were female (85%), which is representative of the Flemish context. In this study, 15 students were bilingual (i.e. Dutch and French).

The virtual learning environment

The VLE is designed in accordance with the 4C/ID-model. The backbone of the VLE is learning tasks related to the teaching of oral interaction in French. Teacher professional knowledge is necessary to solve these learning tasks. The learning tasks are presented in an easy-to-difficult order with diminishing learner support (i.e. instructions or feedback) throughout each task class of the learning environment. Secondly, supportive information is provided. This information is task class-specific. At the beginning of each task class students are directed towards the supportive information (e.g. theoretical frameworks on how to teach oral interaction in French, related vocabulary and grammar, theoretical frameworks on how to prepare a lesson). This information remains available while completing the task class. Thirdly, just-in-time information is presented throughout the learning tasks (e.g. feedback on the exercises). Finally, the learning environment includes part-task practice: mainly grammar and vocabulary and basic pedagogical skills. Links to these exercises are provided throughout the learning tasks. The part-task practices are always related to the learning task the student is working on.

Study design

As aforementioned, for the research design, we aimed to maximize the ecological validity of the study in order to be able to generalize the findings of our study to real-life settings. Accordingly, the administration sessions took place in the students' teacher trainers. The content was delivered by the students' teacher trainers. The content of the virtual learning environment was part of the students' training program and basically replaced a face-to-face session. The first test administrations took place at the beginning of the second semester. During this first administration session students received a pretest (i.e. PCK-, CK- and, PK-test), and the purpose of the online learning environment was explained. After the first administration session students could use the VLE at home for three months. Data of actual use consisting of log files were automatically tracked by the Moodle LMS. During this period the students were encouraged to study the learning content by using the VLE at home. For each module, the participating instructors gave their students a deadline, but still a lot of freedom was given to the students as they could choose how and when to use the VLE within the given time. The students spent an average of 2 hours and 49 minutes on using the VLE. After three months, a second administration session took place. During this session, the students received a posttest and a self-reported questionnaire measuring PU and PEOU. Both were measured simultaneously with the posttest to make sure that the students had sufficient experience with the VLE to have a valid opinion (Turner et al., 2010). During administration session 1 and 2, the participants had 90 minutes to complete the test.

Measurements

Students' performance. To measure teacher professional knowledge of teaching oral interaction French, a quantitative paper-and-pencil instrument was used. The instrument consists of a total number of 212 items and measures pedagogical content knowledge, content knowledge and pedagogical knowledge. The instrument focusses on formal, declarative knowledge. The level of difficulty of the test was aligned with the level that pupils in the Flemish part of Belgium are expected to reach at the end of secondary school, that is level B1 of the Common European Framework of Reference (Evens, Elen & Depaepe, 2017). The students' performance was measured by a pretest-posttest-design, both containing the same content. The pretest was used to measure their prior knowledge. The posttest was used to measure the effectiveness of the intervention on the students' learning outcomes (i.e. students' performance: posttest-pretest). The Cronbach's alpha scores were high (resp., Cronbach's α = .89 for the pretest, and Cronbach's α = .90 for the posttest).

Technology acceptance. The two constructs PU and PEOU were measured on a six-point Likert-type scale ranging from 'strongly disagree' to 'strongly agree'. The PU was measured by six items and PEOU was measured by five items based on the revised item scale of Davis (1989) for PU and PEOU. The statements were adapted to the VLE and translated into Dutch. Scale validity was assessed via confirmatory factor analysis (CFA). The factor reliability was tested by calculating the Cronbach's alpha coefficients (Table 1). All coefficient exceeded .70. After CFA three items (i.e. PU3, PU4, PEOU1) were removed due to very poor factor loadings (< .50 ; Kline, 2013).

Table 2.1.

Confirmatory Factor Analysis

Latent variable	item	Mean *(SD)*	*B*	*α*
PU1	Using the VLE allows me to learn this subject with less effort.	3.96 *(.98)*	.93	.87
PU2	The VLE will make my learning of this subject more effective.	4.27 *(1.10)*	.73	
PU5	Using the VLE makes it easier for me to learn this subject.	4.21 *(1.00)*	.80	
PU6	In general, the VLE helps me to study this subject.	3.97 *(.94)*	.72	
PEOU2	I find it easy to become skillful in working with the VLE.	4.10 *(.77)*	.72	.77
PEOU3	The VLE is easy to use.	4.19 *(.93)*	.85	
PEOU4	The VLE is easy to understand.	4.16 *(.77)*	.55	
PEOU5	In general, I find the VLE user-friendly	4.11 *(.84)*	.50	

Actual use. In order to investigate the impact of PU and PEOU on actual use, actual use was objectively measured using log files that automatically recorded participants' identities and the time they spent on the different components of the 4C/ID-model in the database of the Moodle LMS. Consequently, we were able to capture for each participant the time that he/she spent on using the VLE.

Data Analysis

Firstly, descriptive statistics were used to describe the basic features of the data. Secondly, the correlations between all variables were explored by using a simple correlation matrix and the coefficients of determination were calculated. Thirdly, our research model was fitted in R using the Lavaan package 3.4.0 (Rosseel, 2012). Fourthly, to specify causal relationships between the different variables, parameter estimations were calculated using structural equation modeling (SEM). SEM is a statistical approach for testing hypotheses about the relationships among observed (i.e. actual use and students' learning gain) and latent variables (i.e. PU and PEOU).

2.4. Results

Descriptive statistics, illustrated in Table 2, indicate that there are major differences between students concerning the amount of usage of the VLE (i.e. *SD* = 1 hour, 16 minutes). Looking at the performance variable, we observe an average learning gain of 4.82 (*SD* = 5.22).

Table 2.2.

Descriptive statistics of the manifest variables

	Minimum	*Maximum*	*Mean*	*SD*
Actual use (time)	17	442.00	144.20	84.46
Performance	- 14.82	17.99	4.82	5.22

In the first phase, correlations were investigated. According to the correlation matrix, there is a significant moderate positive relation between PEOU and PU (r = .37, p < .01). Students with higher PEOU are more likely to have a higher PU and vice versa. Furthermore, PU has a significant moderate positive relation with actual use (r = .31, p < .01). Actual use has a moderate significant positive relation (r = .31, p < .01) with performance.

In a second phase SEM-analysis was conducted in order to fit the model and that investigates the relationships between the different variables. SEM was conducted in Lavaan (Rosseel, 2012). For the missing values, a two-stage approach was applied. This approach obtains a saturated maximum likelihood (ML) estimate of the population covariance matrix and then uses this estimate in the complete data ML fitting function to obtain parameter estimates (Savalei & Bentler, 2009). Lavaan converged normally after 24 iterations. The χ^2-test indicates the difference between observed and expected covariance matrixes and should be non-significant. In addition to χ^2 statistics, the root mean squared residual (SRMR), the root mean squared error of approximation (RMSEA), comparative fit index (CFI) and the Tucker-Lewis Index (TLI) were examined. SRMR is the difference between the observed variance and the predicted variance. RMSEA adjusts for the complexity of the model and the size of the sample (Rosseel, 2012). Table 3 summarizes the overall goodness- of fit measures of the model. All values fit well, despite the SRMR, but given the limited sample size (< 200 cases), a value of standardized SRMR < .10 is generally considered adequate. Assessing all measures and considering the above statements, the original structural model was accepted and believed adequate for further analysis (Kline, 2013).

Table 2.3.

Model fit measures

Fit measures	Values	Recommended value
Chi square (χ^2)	40.82; *df* = 33 (*p* = .164)	*p* > .05
RMSEA	.04	< .05
SRMR	.08	< .06
CFI	.96	> .95
TLI	.97	> .95

Figure 2 shows the research model. Research question 1 relates to the influence of PU and PEOU on the actual use of the VLE. First, the influence of PEOU and PU was investigated on actual use. PEOU has a significant influence on PU (β = .47, p < .01). PEOU explains 18% of the variance on PU. Secondly the influence of PU and PEOU on actual use was investigated: PU has a significant influence on actual use (β = .38, p < .01) but PEOU does not (β = .16, p = .15). PU and PEOU account for 18% of the explained variance in actual use. Research question 2 relates to the influence of the actual use of the VLE on students' performance. Actual use has a positive significant effect on students' performance (β = .33, p

< .01) and explains 7% of the variance on performance. These findings suggest that students' PU of the VLE can be used as an indicator of their future use of the VLE and their performance.

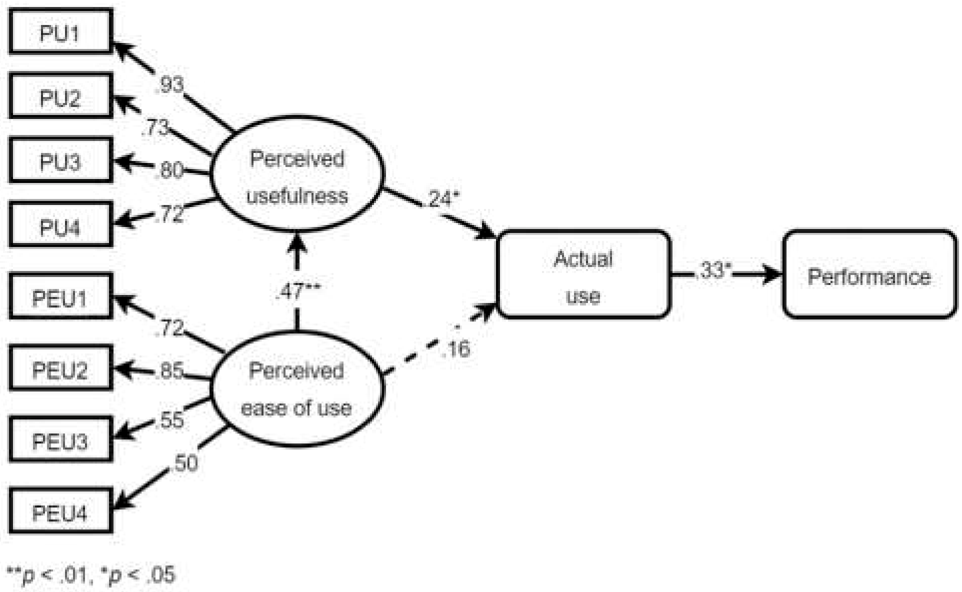


Figure 2.2: Structural equation model with standardized path coefficients

2.5. Discussion

The current study examined whether students' acceptance of a VLE for complex learning is beneficial for students' actual use of the VLE, and students' performance in an ecologically valid context. Accordingly, the current study firstly investigated the direct effect of PU and PEOU on actual use measured objectively by using log files retrieved from the Moodle LMS (i.e. capturing the total time that was spent on using the VLE; RQ1). Results reveal that PU has an influence on actual use. The positive relationship between PU and actual use is in line with former studies. Previous acceptance studies in the context of VLEs also emphasized the importance of PU for the actual use of the VLE (Islam, 2013, Juarez Collazo et al., 2014; van Raaj & Schepers, 2008; Selim, 2003). The results correspond with the literature on the cognitive mediational paradigm which suggests that students should perceive the functionality of the VLE, in order to use it. More specifically, students should recognize that the VLE will be beneficial for their learning process (Clarebout & Elen, 2006). It should be noted that PU explains little variance in usage in the current study. Former studies that studied the relationship between PU and actual use reported stronger relationships (Islam, 2013; van Raaj & Schepers, 2008, Selim, 2003). The major difference in these studies was that actual use was measured by self-reported usage instead of automatically recorded log files (i.e. objective information). As aforementioned in the theoretical framework, self-reported usage exaggerates the causal relationship between independent and dependent variables, which has an influence on the variance explained of the actual use (Benbasat & Barki, 2007). Juarez Collazo et al. (2014) studied the effect of students' acceptance on actual use (i.e. based on log files). Likewise, the current study results indicated that PU had a positive influence on actual use. Nevertheless, in the current study there is no significant relationship between PEOU and actual use, whereas in the study of Juarez Collazo et al. (2014), there is a significant relationship between PEOU and actual use. There is a major difference in the design of the study of Juarez Collazo et al. (2014) that could explain the aforementioned different findings. Students' PU and PEOU was measured before the intervention. In the current study the students' acceptance questionnaire (i.e. PU and PEOU) was filled in after actual use of the VLE together with the posttest. The reason for applying this study design was to make sure that students had sufficient experience with the VLE in order to have a valid opinion about PU and PEOU (Turner et al., 2010). Since PU and PEOU were questioned after the intervention this might have influenced the weak relationship between PEOU and both PU and actual use in this study. This aligns with the study of Venkatesh, Morris, and Davis (2003) as they did not find any direct post-implementation effects of PEOU on BI, only pre-implementation effects. Their findings indicate that as students gain experience with the VLE, PEOU is overshadowed by other factors (Schepers & Wetzels, 2007). This could be an explanation for the differences in the relationships between PEOU and PU and actual use in the study of Juarez Collazo et al. (2014) and the present study.

Secondly (i.e. RQ2) the current study investigated the effect of actual use on students' performance (i.e. posttest-pretest). Results reveal that actual use has a significant effect on students' performance. Findings are in line with the study of Juarez Collazo et al. (2014), who also found that actual use had a positive effect on students' performance (posttest-pretest). The strength of the relationships is similar in both studies, which supports the view that student engagement (i.e. time-on-task) in learning exercises has an important influence on academic achievement (Revere & Kovach, 2011; Slavin, 2003). Nevertheless, in the study of Juarez Collazo et al. (2014), the total variance explained of performance was higher than in the current study. A possible explanation is that students' performance could also be influenced by other courses. Since the study was ecologically valid, other courses containing CK, PCK, and PK were taught during the study and accordingly, this could have had an influence on students' learning gain instead of using the VLE.

Limitations and future directions

The current study has some important limitations. Firstly, the observation that a lot of variance of actual use in the current study remains unexplained suggests the need for additional research in incorporating potential unmeasured variables in the current study. Firstly, since the study took place in its actual context (i.e. teacher training institute), the students were encouraged by their instructor to use the VLE. In this study, the influence of the instructor on the students' actual use was not measured (i.e. subjective norm). It should be measured as an external variable both influencing PU and actual use. Venkatesh and Davis (2002) hypothesized that subjective norm influenced both PU and BI, assuming that learners often choose to perform an action when one or more important referents say they should. Additionally, Legris et al. (2003) emphasized the need to include human and social change process variables (e.g. subjective norm). Moreover, the strong relationship between subjective norm on PU was once more confirmed by a meta-analysis performed by Schepers and Wetzels (2007). Furthermore, they showed that using a student sample seriously affected these relationships, since students have a stronger tendency to comply with authority. These findings raise the need for further research to examine the influence of the instructor on PU and actual use. A second important remark is that the VLE is based on the First Principles of Instruction of Merrill (2002). Since the VLE is based on the First Principles of Instruction, this should have a positive influence on students' performance (Sarfo & Elen, 2011). Frick et al. (2010) studied the effect of using the First Principles of Instruction and found that the likelihood of a high level of student mastery of course objectives (i.e. student performance) is about five times greater than the likelihood when neither First Principles of Instructions occurred.

Additionally, a well-considered instructional design could also influence students' perceptions towards the online courses (Song et al., 2003; Rienties & Toetenel, 2016). Accordingly, it would be interesting for further research to complement the TAM framework with external indicators to investigate the effect of the perceived quality of the instructional design on students' perceptions (i.e. PU and PEOU), use and students' performance. Thirdly, even though Merrill (2002) claims that incorporating the First Principles of Instruction in a learning environment promotes student engagement, the importance of the degree of interaction in a VLE cannot be overlooked. Former studies indicated that students' perceptions towards VLEs are tied to the degree of interaction with the instructor or peers (Jaggers & Xu, 2016). Additionally, not only the pedagogical design seems to influence students' perceptions, but also the technology used to deliver the instructional material. In the current study we did not take into account the effect of the technology used, meanwhile former research indicated that effective use of IT in delivering e-learning based components of a course is of critical importance to the success and students acceptance of online learning (Selim, 2007; Younis Alsabawy, Cater-Steel & Soar, 2016). Fourthly, prior research revealed that differences in tool use between students affected performance significantly (Lust et al., 2012). The current study examined the amount of time spent on the VLE and related this to students' final performance. This log-indicator is too broad, since it gives little insight into why using a VLE existing of four different components containing different information (or different formats of information: summary, theory, part-task-practice), relates to learning. Specifying log-indicator or combining multiple log-indicators can give a more detailed picture of how the different components of a 4C/ID-based VLE support learning.

2.6. Conclusion

In conclusion, the objective of this study was to examine the strength of the path coefficients in a highly ecological valid study design. More specifically, we wanted to investigate whether students' acceptance of a VLE influences objectively measured use and performance. With an instructional design based on the 4C/ID-model, the educational objective, context, and students' needs were highly considered (Elen, 2004). Findings suggest that students' PU of the VLE can be used as an indicator of their future use of the VLE in an ecological context. Furthermore, results show that the actual use of the VLE has a positive influence on the students' performance

Chapter Three

The perceived instructional quality

The text of this chapter was published as: Larmuseau, C., Desmet, P., & Depaepe, F. (2018). Perceptions of instructional quality: impact on acceptance and use of an online learning environment. *Interactive Learning Environments, 7*, 953-964. https://doi.org/10.1080/10494820.2018.1509

3. The perceived instructional quality of a 4C/ID-based online course

3.1. Introduction

Online learning is increasingly important in higher education since it has many benefits, such as reduced education costs and flexible accessibility of education. Given the growing use of online learning environments, it is crucial to define the factors that influence the effectiveness of the online learning environments. Many researchers claim that instructional quality can contribute to the effectiveness of online courses (Czerkawski & Lyman, 2016; Terras & Ramsay, 2015). Accordingly, it is important to define quality indicators for designing online learning environments that promote students' engagement. Merrill (2002) claims that the instructional design can promote students' engagement and learning outcomes when (1) learners are engaged in real-world problems, (2) existing knowledge is activated, (3) new knowledge is demonstrated, (4) new knowledge is applied and (5) new knowledge is integrated into the learner's world. These fundamental principles known as Merrill's First Principles of Instruction, underpin all contemporary instructional design models and theories and can be implemented in any delivery system using any instructional architecture (Margaryan, Bianco & Littlejohn, 2015; Merrill, 2002).

Nevertheless, evidence from research indicates that applying the above-mentioned design principles is by no means a guarantee of success. As the instructional design and students' characteristics are largely interrelated, students' perceptions of the instructional quality are crucial to ensure its effectiveness (Martens et al., 2007; Terras & Ramsay, 2015). By way of illustration, the question is not only if a task is authentic, but whether it is perceived as such by the students (Martens, et al., 2007). Former research already indicated the influence of students' perceptions on course effectiveness. Technology acceptance studies indicated the important influence of the perceived instructional quality on students' acceptance of an online learning environment (Lee, Yoon & Lee, 2009; Liaw & Huang, 2013; Yang, Shao, Liu & Liu, 2017). Moreover, results revealed that students' acceptance can influence the quantity of use in online courses (Juarez Collazo et al., 2014). Additionally, the study of Frick et al. (2010) revealed that integrating the First Principles of Instruction of Merrill (2002) in the course design is related to high levels of mastery of course objectives. Although, the aforementioned studies explored students' perceptions of the quality of the instructional design and students' technology acceptance, research on these perceptions related to the quality of the use remains scarce. Meanwhile, former research indicates that the quantity of use does not necessarily mean that students used the online learning environment adequately (i.e. in line with the instructional intentions; Juarez Collazo, Elen & Clarebout, 2015). Moreover, little information is given about the instructional design of the studied online learning environments (Lee et al., 2009; Liaw & Huang, 2013; Yang, Shao, Liu, & Liu, 2017). As a result, we can only formulate limited recommendations for instructional designers on what quality indicators influence acceptance and use of an online learning environment. To deal with current limitations, this study investigates students' perceptions and use of an online learning environment based on the guidelines of the 4C/ID-model from van Merriënboer (1997). The 4C/ID-based online course incorporates the First Principles of Instruction of Merrill (2002).

The aim of this study is twofold. The first aim is to investigate the influence of the perceived instructional quality on students' acceptance. A second aim is to investigate the impact of technology acceptance and the perceived instructional quality on the quantity and quality of use. In the current study, perceived instructional quality is based on the First Principles of Instructional of Merrill (2002), and technology acceptance is based on perceived usefulness and PEOU retrieved from the TAM (Davis, 1989). Quantity of use is based on the students' course activity which is automatically retrieved from the Moodle LMS and the quality of use is based on students' course performance (e.g. the quality of students' answers to different tasks). Overall, finding instructional quality factors of an online learning environment that influences students' acceptance and students' quantity of use and quality of use can be instrumental information to support developers of online learning environments when designing their online courses.

3.2. Theoretical background

The quality of the instructional design

Many current instructional design models indicate that effective learning can be promoted when online learning environments are problem-centered and involve the students in four distinct phases of learning (a) activation of prior knowledge, (b) demonstration of knowledge, (c) application of knowledge, and (d) integration of knowledge into the learners' world (Jonassen, 1999; Merrill, 2002). Accordingly, Merrill (2002) elaborated five prescriptive design principles for problem-centered instruction for each of the four instructional phases, namely, the First Principles of instruction: (1) authentic problems: learning is promoted when learners are engaged in solving real-world problems, (2) activation: learning is promoted when existing knowledge is activated, (3) demonstration: learning is promoted when new knowledge is demonstrated to the learner, (4) application: learning is promoted when new knowledge is applied by the learners, (5) integration: learning is promoted when new knowledge is integrated into the learners' world.

An instructional design model that incorporates these design principles, is the 4C/ID-model developed by van Merriënboer (1997). The 4C/ID-model involves four interrelated components: learning tasks, part-task practice, supportive and just-in-time information (van Merriënboer, 1997; van Merriënboer & Kirschner, 2018). The backbone of the instructional design is concrete, authentic, problem-based, whole-task experiences learning tasks i.e. authentic problems that are grouped in a task class. Activation of prior knowledge is promoted by consulting support. More specifically, just-in-time information encourages learners to recall recurrent aspects of a task that can be used to organize the new knowledge, and supportive information is provided containing nonrecurring aspects of the learning task to promote elaboration. The learning tasks start with a worked example i.e. demonstration, that not only confront the students with the desired goals but also show the learner how to deal with a problem (van Merriënboer & Kirschner, 2018). The instructional design also stresses the application of knowledge and skills to solve problems. These problems are embedded in the learning tasks and part-task-practices, consistent with the learning objectives. Integration is at the center of the 4C/ID-model, since the whole-task practice design leads the student towards a real-world task (Merrill, 2002). The learning tasks are sequenced based on their complexity. Merrill (2002) argues that if one of these First Principles of Instruction is lacking, this can negatively impact students' performance. Frick et al. (2010) conducted a study in which they investigated the association between students' perceived quality of the course design based on the First Principles of Instruction and students' course achievement. Results indicated that students who agreed that these First Principles of Instruction were presented were more likely to achieve high levels of mastery of course objectives. Moreover, former research also indicated that quality indicators of the instructional design of an online learning environment can influence students' acceptance of the online learning environment (Lee et al., 2009; Liaw & Huang, 2013; Yang et al., 2017). An overview of this research is provided in the next section.

Technology acceptance

A model that is frequently used to explain why an individual accepts or rejects IT from a social psychological perspective is the technology acceptance model (TAM) proposed by Davis (1989). According to the study of Šumak et al. (2011) who systematically reviewed existing knowledge in the field of online learning acceptance, TAM is the most common ground theory in online learning acceptance literature. TAM postulates that both PU and PEOU indicate students' technology acceptance. PU is defined as the degree to which the user believes that using the IT will improve his or her learning performance, whereas PEOU is the degree to which the user believes that using the IT will be user-friendly. According to TAM, PEOU has an important influence on perceived usefulness (Davis, 1989).

Former studies have already used TAM as a baseline, to investigate quality factors of the instructional design that influence students' acceptance of online learning environments (Lee et al., 2009; Liaw & Huang, 2013; Yang et al., 2017). For instance, Liaw, and Huang (2013) conducted SEM to investigate the influence of the quality indicators of the instructional design (i.e. self-reported questionnaire about the interactivity) on PU of online learning environments. Results of 191 university students reveal that PU is influenced by the level of interactivity of online learning environments. Similarly, Lee et al. (2009) conducted a study where they investigated the influence of four perceived quality indicators, such as, instructor characteristics (e.g. clear instructions), teaching materials (e.g. fit with the learning objectives), design of learning content (e.g. a variety of learning content) and playfulness (e.g. improves creativity), on students' acceptance of online learning based on regression analysis. Results of 250 undergraduate students indicate that instructor characteristics and teaching materials are positively related to PU. Furthermore, they observed that the design of learning content is positively related to PEOU. Moreover, Yang et al. (2017) developed a model for investigating quality indicators of the course design that influences students' acceptance of Massive Open Online Courses (MOOCs). The quality indicators where system quality (i.e. integration of system functions and reliability of system operations), course quality (i.e. source and content of the information) and service quality (i.e. support delivered by the MOOC's service provider). These quality indicators were integrated into the model as predictors of students' acceptance. Based on an SEM-analysis of 294 respondents, results indicated that system quality has a significant impact on PEOU and, course quality and service quality are significant antecedents of PU. Results of the previous studies indicate that the perceived quality of the instructional design can have an influence on students' acceptance of online learning environments. Moreover, quality indicators of the course design seem to influence both PU and PEOU, whereas meaningful content of the online learning environment seems to have the tendency to exert more influence on PU.

Given that students' acceptance alone does not guarantee the effectiveness of the online learning environment, former studies have extended TAM, by adding the quantity of use as the dependent variable of perceived usefulness and PEOU (King & He, 2006; Schepers & Wetzels, 2007). The study of Juarez Collazo (2014) investigated the effect of students' acceptance on the quantity of use (i.e. time spent in the online learning environment) and quality of use (i.e. analysis of the answers) and found a positive effect of PU and PEOU on the quantity of use, but none on the quality of use. Larmuseau et al. (2018) investigated the effect of students' acceptance of actual use (i.e. time spent) in an ecologically valid context and found a positive impact of PU on the quantity of use. Both studies indicate that students' acceptance can influence the quantity of use of an online learning environment. Summarized, the perceived instructional quality of the online learning environment can have a positive influence on students' acceptance. Moreover, the perceived instructional quality and students' acceptance can have an influence on the quantity of use (Juarez Collazo et al., 2014; Lee et al., 2009). Furthermore, results also indicate that the perceived quality of the instructional design can promote student engagement and course performance (Frick et al., 2010; Martens et al., 2007).

Therefore in the current study, the first aim is to investigate the influence of perceived instructional quality operationalized in terms of the First Principles of Instruction on students' acceptance. As we are interested in the impact of students' perceptions on both quantity and quality of use, a second aim is to investigate the impact of students' PU, PEOU and the perceived instructional quality on the quantity (i.e. course activity) and quality (i.e. course performance) on use. It should be noted that former studies gave little information about the instructional design of the investigated online learning environments. Therefore, major attention was given to the instructional design of the online learning environment as it was systematically designed according to the 4C/ID-guidelines. Furthermore, the online course incorporates the First Principles of Instruction in order to have insight on how these instructional design principles influence students' PU, PEOU, quantity, and quality of use. A theoretical research model is proposed as shown in *Figure 1* containing all variables in order to elucidate the relationships among these variables.

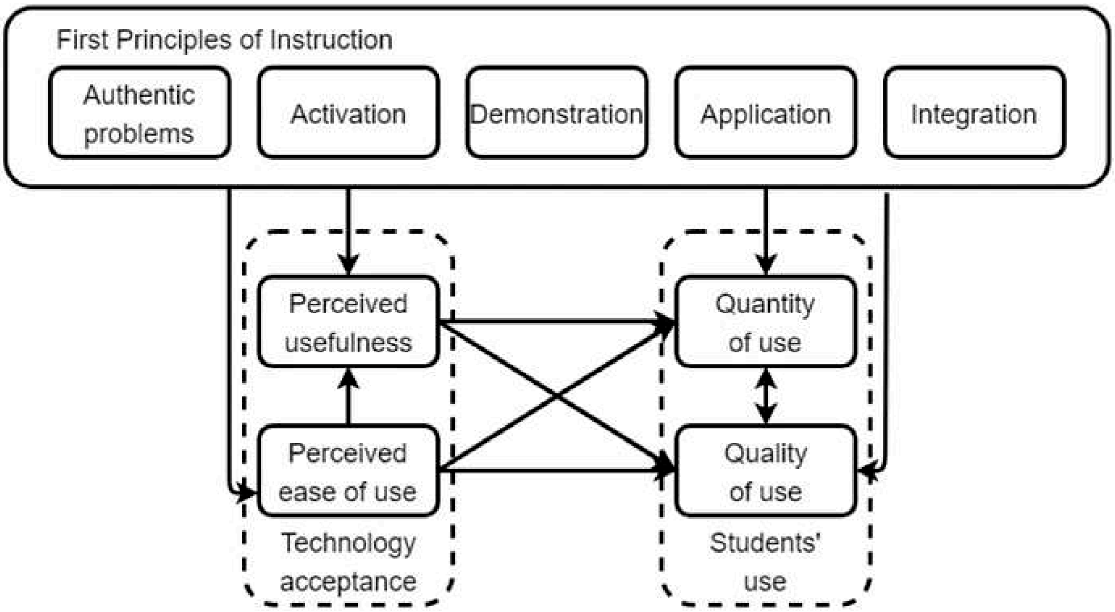


Figure 3.1.: The research model

3.3. Method

Participants

The study took place in the Flemish part of Belgium. The participants were 161 first year Psychology and Educational Science university students. The majority of the students were female (91%). The average participant was 20 years old (*SD* = 2.92) which is representative for the student population in Educational Sciences (Conger & Dickson, 2017). Participation to research is part of the students' training program. Before answering the questions, informed consent was obtained from all individual participants included in the study.

Study design

The first administration session started with an introduction to the learning environment. After the exploration phase, the students were asked to use the learning environment at home for two weeks. The second administration session took place after those two weeks and consisted of a self-reported questionnaire about the quality of the instructional design (i.e. authenticity, activation, demonstration, application, and integration) and a self-reported questionnaire with the TAM constructs (i.e. PU, PEOU). The perceived quality of the instructional design and the TAM constructs were questioned after students used the online learning environment in order to make sure students had sufficient experiences with the learning environment to have a valid opinion.

The online learning environment

The online learning environment in the present study teaches French as a foreign language. The level of difficulty was aligned with the level that pupils in the Flemish part of Belgium are expected to reach at the end of the secondary school (i.e. level B1 of the Common European Framework of Reference; Evens, Elen & Depaepe, 2017). The online learning environment is developed along the lines of the 4C/ID-model (Van Merriënboer, 1997) and, accordingly, aligns with the First Principles of Instruction (Merrill, 2002). The problems formulated by the learning tasks are based on authentic situations (e.g. ordering food in a restaurant) in easy-to-difficult order with diminishing learner support. Each learning task starts with a worked example (e.g. video of someone explaining the way) combined with clearly defined course objectives (i.e. demonstration). During the learning tasks, students have a lot of opportunities to practice what they've learned i.e. application. Supportive information (e.g. explanation of grammar) is provided which is the bridge between what students already know and what they still have to learn in order to solve the learning tasks (i.e. activation). Support is also provided via procedural information. At the beginning of each task, students can click – through pop-ups – on additional procedural information, which contains just-in-time support to complete a particular learning task. The pop-ups contain, for instance, an overview of grammar or vocabulary needed for the specific exercise within the learning task. The part-task practices (e.g. drill-and practice exercises about verbs) are always related to the learning task the student is working on. The learning tasks are related to real-world contexts i.e. integration. The learning tasks and part-task practices were automatically corrected and adaptive feedback was generated. Interaction with the instructor was possible via the discussion forum or with fellow students via the chatbox.

Measurements

Quantity and quality of use. Information on students' use of the online learning environment was collected through logging students' activity (e.g. registration of views, quiz attempts, tasks submitted) during two weeks (i.e. after measurement moment 1 and until measurement moment 2). Students' quality of use basically refers to optimal use of the online course and was measured on the basis of students' course performance (e.g. quality of the provided answers). Students' course performance is the average score of the students in the online course (i.e. scores on learning tasks and part-task practices). Results of the learning tasks and part-task practices were automatically rated by Moodle LMS. The learning tasks with open-ended questions were corrected by the instructor.

Perceived instructional quality. Students were asked to fill in the Teaching and Learning Questionnaire (TALQ). This is a self-reported questionnaire about the First Principles of Instruction originally developed by Frick, Chadha, Watson, Wang, and Green (2009) to evaluate the instructional design of courses. The five constructs (i.e. authenticity, activation, demonstration, application, and integration) were questioned on a 7-point Likert scale. As we are interested in the combined effect of the First Principles of Instruction, second-order confirmatory factor analysis (*CFA*) was conducted (*Figure 2*). This is a statistical method used to confirm that the underlying sub-constructs load into the construct perceived quality of the instructional design (Black, Yang, Beitra & McCaffrey, 2015). Construct validity of the construct perceived instructional quality and the sub-constructs was investigated by examining convergent validity and reliability (Table 1). Composite reliability (*CR*) was determined in order to investigate how well the items for one construct correlate with each other. The average variance explained (*AVE*) was determined in order to investigate the level of variance captured by a construct versus the level due to measurement error. Reliability was determined through Cronbach's α in order to investigate the overall consistency of the constructs (Schreiber, Nora, Stage, Barlow & King, 2006). Items with poor factor loadings (< .50) were removed. Application and integration were merged into one construct since the constructs were highly correlated. After revision all constructs were significant and all factor loadings exceeded .50. The average variance explained (AVE) ranged from .35 to .76, and accordingly, not all constructs reached the threshold of .50. Nevertheless, the AVE of the construct perceived instructional quality is .54. CR ranged from .62 to .91 with .82 for the construct perceived instructional quality. The Cronbach's α for all sub-constructs ranged from .60 to .91, and was .84 for perceived instructional quality. Subsequently, for both Cronbach's α and CR the construct perceived instructional quality exceeded the threshold of .70 (Cuieford, 1965; Rosseel, 2012).

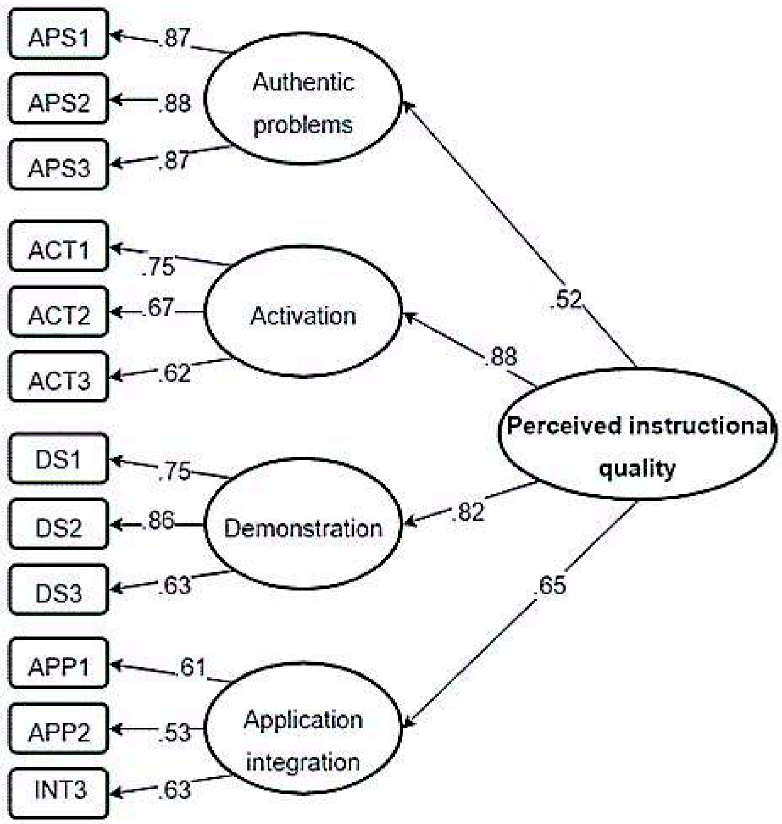


Figure 3.2. Factor loadings of perceived instructional quality

Technology acceptance. The constructs PU and PEOU (i.e. 4 items) were retrieved from TAM. The Cronbach's α was measured and CFA was conducted for each construct. All factor loadings were significant (*Figure 3*). AVE for PU was .56 and .65 for PEOU (i.e. should exceed .50). Cronbach's α for PU was .74 and for PEOU .87, suggesting that the constructs are highly reliable (Cuieford, 1965; Rosseel, 2012). An overview of the survey items can be found in Table 1.

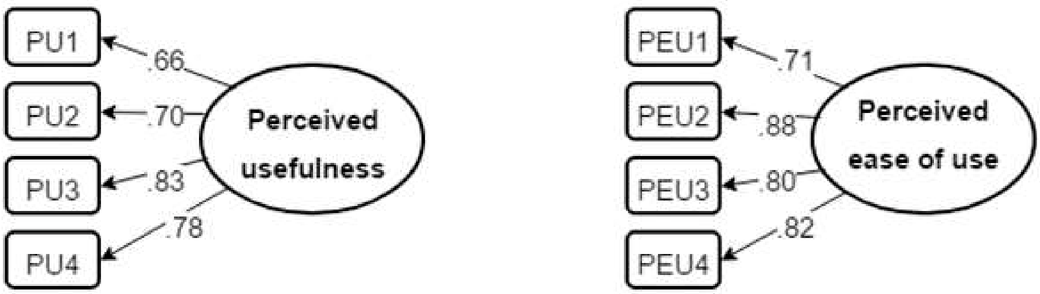


Figure 3.3. Factor loadings of PU and PEOU

Table 3.1.

Survey items, reliability and convergent validity of the constructs

Construct	Measurement items	AVE	A	CR
Authentic problems	APS1: I performed a series of increasingly complex authentic tasks in this course. APS2: I solved authentic problems of completed authentic tasks in this course. APS3: I solved a variety of authentic problems that were organized from simple to complex.	.76	.91	.91
Activation	ACT1: I engaged in experiences that subsequently helped me learn ideas or skills that were new and unfamiliar to me. ACT3: The course design provided a learning structure that helped me mentally organize new knowledge and skills. ACT4: In this course I was able to connect my past experience with new knowledge and skills.	.46	.71	.71
Demon-stration	DS1: The course demonstrated what I was expected to learn. DS2: The course provided examples and counter-examples of concepts that I was expected to learn. DS3: The course provided alternative ways of understanding the same ideas or skills.	.62	.83	.83

Application /integration	APPLI1: I had opportunities to practice or try out what I learned in this course. APPLI2: I had opportunities to explore how I could personally use what I have learned. INT3: I was able to demonstrate what I learned in this course.	.35	.60	.62
PU	PU1: This course gives me more control to learn the content. PU2: This course enhances my effectiveness of learning the content. PU3: This course makes it easier for me to learn the content. PU4: In general, this course will help me to learn the content.	.56	.83	.83
PEOU	PEU1: It's easy for me to become skillful in working with the course. PEU2: The course is clear and understandable. PEU3: The course is easy to use. PEU4: In general, I find the course user-friendly.	.65	.87	.88

Analysis

Firstly, descriptive analyses such as means, standard deviations, and correlation analyses were conducted for the variables, that is, perceived quality of the instructional design, PU, PEOU, quantity (i.e. course activity), and quality of use (i.e. course performance). Secondly, in order to gain insight into (1) the influence of the perceived quality of the instructional design on students' acceptance, (2) the influence of students' acceptance and, students' perceived instructional quality on the quantity and quality of use, SEM was conducted in R using the Lavaan package 3.4.0. We employed the maximum likelihood parameter estimates (MLM -chi square) with standard errors and a mean-adjusted chi-square test statistic that is robust to non-normality (i.e. Satorra-Bentler chi-square). This technique was applied because of the minimal demands on data assumptions, such as multivariate normality assumptions (Rosseel, 2012). Missing data was found for some of the self-reported measures, such as the perceived instructional quality, PU, and PEOU (1.9%).

3.4. Results

Preliminary analysis

An overview of descriptive statistics can be found in Table 2. Nearly all students consulted the learning tasks (N = 158). The average time spent on using the online learning was 66 minutes (SD = 27.34).

Table 3.2.

Descriptive statistics

Variables	*N*	*min*	*max*	*Mean*	*SD*
Instructional quality	158	1.42	6.25	4.72	.78
PU	158	1.75	6.75	4.31	.89
PEOU	158	1.50	7.00	5.55	.84
Quantity of use	158	23	326	109.81	47.22
Quality of use	158	18.99	100	73.59	16.14

Table 3 gives an overview of the correlations. Correlations reveal that PU and PEOU are positively correlated. Perceived instructional quality is significantly positively related to PU and PEOU. Perceived instructional quality and PEOU are positively significantly related to the quality of use, but not to the quantity of use. PU is not correlated with the quantity and quality of use.

Table 3.3.

Correlations

	1	2	4	5	6
1. Instructional quality	1				
2. PU	.34**	1			
4. PEOU	.35**	.21**	1		
5. Quantity of use	.13	.03	.09	1	
6. Quality of use	.28**	.09	.18*	.43**	1

*** correlation is significant at the .01 level/* correlation is significant at the .05 level*

Structural equation modeling

SEM analysis was conducted to (1) investigate the influence of the perceived instructional quality on students' acceptance of the online learning environment, (2) to investigate the influence of students' perceived instructional quality and students' acceptance on the quantity and quality of use. The hypothesized model, provided an adequate fit to the given data [χ^2/df = 274.43/189.7= 1.4; SRMR = .07, RMSEA = .05, CFI = .94, TLI = .92]. The χ^2-test indicates the difference between observed and expected covariance matrixes. The normed χ^2-test should have a value smaller than 2.0. SRMR is the difference between the observed variance and the predicted variance. A value of less than .08 is considered a moderate fit. RMSEA adjusts for the complexity of the model and the size of the sample. The value of RMSEA for acceptance is .6. A value of CFI and TLI between 0.90 and 0.95 is acceptable. Assessing all measures and considering the above statements, the original structural model was accepted and believed adequate for further analysis (Schreiber et al., 2006). *Figure 4* gives an overview of the standardized path coefficients. The solid lines show significant relationships. Results reveal that the perceived instructional quality has a significantly positive influence on both PU (β=.55, p <.01, R^2 =.22), and PEOU (β =.56, p <.01, R^2 =.23). PEOU has no significant influence on PU. Regarding the influence of students' acceptance on students' quantity of quality of use, results reveal that PU and PEOU have no influence on the quantity and quality of use. Furthermore, the perceived instructional quality has a significantly positive influence on the quality of use (β =.37, p <.05, R^2 =.12), but not on the quantity of use. In conclusion, results indicate that the pedagogical aspects of the online learning environment have a positive significant influence on students' acceptance, and on the quality of use of the online learning environment.

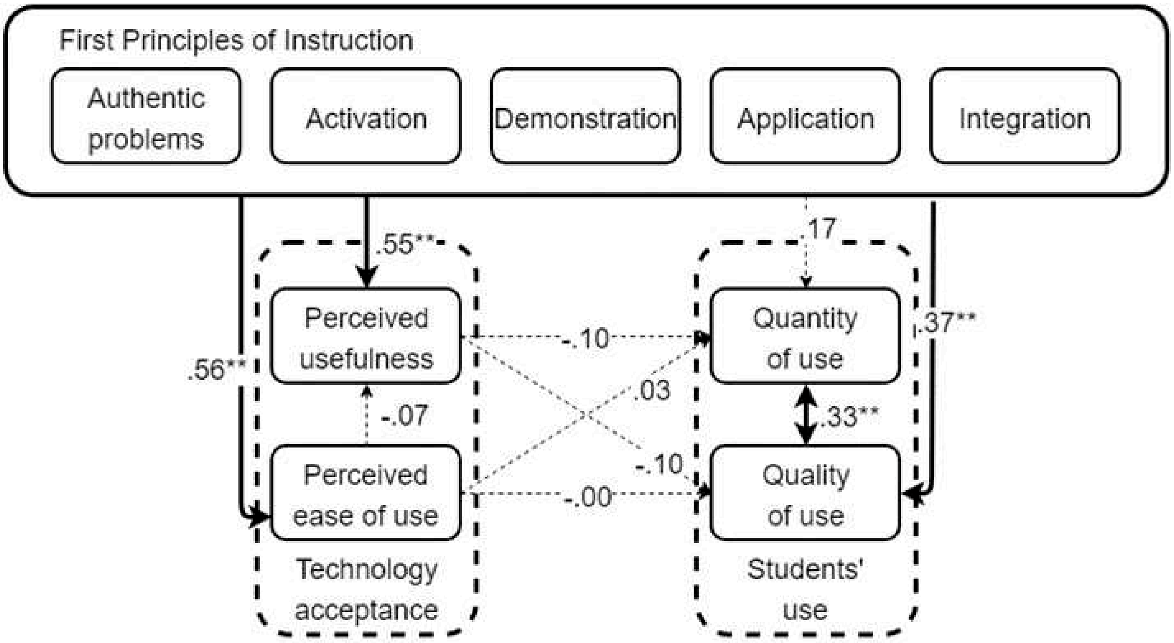


Figure 3.4. Structural model with standardized path coefficient

The research model

This paper attempted to firstly analyze the influence of perceived instructional quality on students' acceptance of a systematically designed online learning environment incorporating the First Principles of Instruction (Merrill, 2002). Secondly, it analyzed the influence of students' acceptance and perceived instructional quality on the quantity and the quality of use. Results reveal that the perceived instructional quality influences both PU and PEOU. Results therefore indicate that content-related aspects influence perceptions of usefulness and user-friendliness. These findings are in line with the study of Lee et al. (2009) as this study found a direct effect of instructor characteristics (e.g. clear instructions) and teaching materials (e.g. fit with the learning objectives) on PU and a direct effect of the design of the learning content (e.g. variety of the learning content) on PEOU. Moreover, the study of Yang et al. (2017) found that perceived course quality and service quality had a direct influence on PU and that perceived system quality had a direct effect on PEOU in a MOOC. In contrast to the original TAM model, the current study reveals that PEOU has no influence on PU (King & He, 2006; Šumak, et al., 2011). Accordingly, this finding indicates that the perceived quality of the instructional design has more impact on PU, than

PEOU. More specifically, this could indicate that pedagogical aspects of the course design seems to exert more influence on the PU, than the user-friendliness.

Additionally, the influence of students' acceptance on the quantity and quality of use was investigated. In this study, no influence was found of students' PU and PEOU on the quantity and quality of use. This is in contrast with the studies of Juarez Collazo et al. (2014) and Larmuseau et al. (2018). Both studies indicated a positive influence of PU on the quantity of use (i.e. time spent). The major difference in the study of Larmuseau et al. (2018) was that this study was highly ecologically valid as the online course was part of the students' training program. In contrast, the current study had an experimental design which implies that the quantity and quality of use were probably very dependent on students' intrinsic motivation to learn French. Conversely, the study of Juarez Collazo et al. (2014) had a strong experimental design as students had to use an online learning environment while being supervised in a computer room. Accordingly, results might be influenced (e.g. social desirability response; compulsory use).

Finally, the current study investigated the direct link between students' perceived instructional quality and their quantity and quality of use of the learning environment. Results revealed that students' perceived instructional quality influences the quality of use, but not the quantity of use. These findings could be related to the findings of Frick et al. (2010) who indicated that students' performance is related to the integration of the First Principles of Instruction in the course design. Accordingly, findings indicate that instructors can improve the effectiveness of their online courses by implementing the First Principles of Instruction. Summarized the perceived instructional quality based on the First Principles of Instruction seems to influence both PU and PEOU. Additionally, perceived instructional quality has a positive significant influence on the quality of use.

Limitations

Despite the merits of the study in terms of emphasizing the important role of students' perceived quality of the instructional design, some important limitations should be mentioned. Although the perceived quality was based on the First Principles of Instruction, an important quality indicator, especially for online learning, namely the degree of interaction was overlooked (Liaw & Huang, 2013). As the online learning environment in the current study disposed of a chatbox (i.e. contacting fellow students) and a discussion forum (e.g. contact with the instructor), this could also have been an important indicator for students' quantity and/or quality of use.

A second limitation concerning methodological aspects, is that the self-reported questionnaire measuring the perceived instructional quality (i.e. TALQ) was largely revised due to low factor loadings. Possibly, this major revision was due to the fact that the questionnaire was originally developed to evaluate traditional face-to-face courses (Frick et al., 2009). Accordingly, some items were less adapted to the specific online learning context. Thirdly, it should be mentioned that the sample size of this study was rather low, given the complexity of the model (i.e. amount of latent constructs), which could undermine statistical power. Nevertheless, given the appropriate factor loadings, the small amount of missing values and taking the rules of thumb (i.e. minimum sample size of 100) into account, we assumed that SEM was suitable for this study (Wolf, Harrington, Clark & Miller, 2013). Fourthly, the sample consisted mainly out of women. As the sample consisted mainly of women, we should be careful when generalizing these results. Future research in this area should also include male students to see whether the results are similar. Also the topic of the online course could have an influence on the results, since former studies indicated that female students are generally more motivated to learn French as a foreign language (McIntyre, Baker, Clément & Donovan, 2003). A fifth important limitation of the study design is the perceived instructional quality and the TAM constructs were measured after students used the online learning environment. The major reason for this event was to make sure that students had sufficient experience with the online learning environment. Nevertheless, this implies that we should be careful when making claims of causality. Studies that use similar research models must verify the current relationships in order to be able to generalize them. Additionally, it would be interesting to conduct future studies in a more ecologically valid setting as this could influence the relationship between technology acceptance and the quantity of use (Larmuseau et al., 2018).

3.5. Conclusion

To summarize, findings reveal an impact of perceived instructional quality based on the First Principles of Instruction (Merrill, 2002) on students' technology acceptance. Moreover, the perceived instructional quality seems to exert more influence on PU than the user-friendliness of the online learning environment, and therefore plays a more important role in students' acceptance of the online learning environment. Nevertheless, results indicate that students' acceptance does not influence the quantity and quality of use. By contrast, current study emphasizes the important positive influence of the perceived instructional quality on the quality of use in terms of course performance. Accordingly, this study emphasizes the importance of a well-considered instructional design on students' acceptance and quality of use. Future studies should conduct similar studies in ecologically valid settings, with larger sample sizes in order to verify the direction and strength of these relationships.

Chapter Four

Cognitive and motivational characteristics

The text of this chapter was published as: Larmuseau, C., Elen, J., & Depaepe, F. (2018). *The influence of students' cognitive and motivational characteristics on students' use of a 4C/ID-based online learning environment and their learning gain.* In Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18), Sydney, Australia. 171–180. https://doi.org/10.1145/3170358.3170363

4. The influence of cognitive and motivational characteristics on differences in use

4.1. Introduction

Educational research shows that the effectiveness of an online learning environment depends a great deal on its instructional design (Van Laer & Elen, 2017). Moreover, the instructional design can influence the behavior and performance of students in online learning environments (Rienties & Toetenel, 2016). An instructional design model that is acknowledged as one of the most effective instructional design models for designing effective learning environments is the 4C/ID-model (Cook & McDonald, 2008). A 4C/ID-based online learning environment contains four different components, namely, learning tasks, part-task practice, supportive and procedural information, and confronts the learner with the need to assess his or her own performance. Accordingly, on the learner's initiative, additional tasks and support can be selected which implies that learners can choose their own learning trajectory to a smaller or larger degree. Consequently, we expect that one student may quickly proceed from learning task to learning task, while another learner might select part-task practice or consult supportive information (van Merriënboer & Sluijsmans, 2009). Former research findings indicate that students' use of different components in an online learning environment can differ based on cognitive and motivational characteristics (Greene & Azevedo, 2007; Jiang, Elen & Clarebout, 2009; Rienties, Tempelaar, Van den Bossche, Gijselaers & Segers, 2009).

Therefore, the first aim of this study is to investigate the influence of students' cognitive (i.e. prior knowledge) and motivational (i.e. task value and self-efficacy) characteristics on students' use of the four components of a 4C/ID-based online learning environment. Furthermore, the provision of different components can stimulate students' performance since students can consult various support during their learning (Lust et al., 2013). By looking at the four components separately, more insight can be gained in how students' use of a specific component contributes to students' learning gain. Therefore, the second aim of this study is to measure how students' use of the four components of the 4C/ID-model, influences students' learning gain, taking into account their cognitive and motivational characteristics. In order to achieve both aims, a theoretical structural model is suggested that integrates students' cognitive (i.e. prior knowledge) and motivational (i.e. task value and self-efficacy) characteristics, the four components of the 4C/ID-model and course performance in order to elucidate the relationships among these variables.

4.2. Theoretical background

Instructional design

A learning environment should promote constructive, active, cumulative and self-directed learning (Elen, 2015; Greene & Azevedo, 2007). More specifically, well-designed tasks should stimulate students to integrate required skills, knowledge and attitudes and transfer complex cognitive skills to real-world contexts.

An instructional design model that stresses integration and transfer of learning is the 4C/ID-model elaborated by van Merriënboer, Clark & de Croock (2002). The 4C/ID-model is acknowledged as one of the most effective instructional design models for designing effective learning environments that facilitate the acquisition of integrated sets of knowledge, attitudes and skills (Cook & McDonald, 2008). When designing an online learning environment the first question to ask is how to develop learning tasks in order to reach the educational objectives and how to organize and sequence these learning tasks (Elen, 2015). The basic concept of the 4C/ID-model is that learning environments can be described in terms of four interrelated blueprint components: (1) learning tasks, (2) part-task practice, (3) supportive and (4) just-in-time information (van Merriënboer et al., 2002, 2003). In a 4C/ID-based learning environment, the learning tasks are the backbone of the design process and are concrete, authentic, problem-based, whole-task experiences. The design of the learning tasks, allows for simultaneous practice of domain-specific knowledge and cognitive strategies. Learning tasks are grouped in task classes and sequenced based on their degree of difficulty in order to prevent cognitive overload for the learners, as this could hamper learning and performance (Greene & Azevedo, 2007; Merrill, 2002). Learners can encounter difficulties while solving learning tasks, consequently adapted learner support should be provided (Elen, 2015). Support is provided in two distinct manners, that is, supportive and procedural information (van Merriënboer et al., 2002, 2003). Supportive information is basically, the theory and therefore supports the learning and performance of the non-recurrent, problem-solving and reasoning aspects of learning tasks. It helps learners to link the presented information to existing schemata, that is, to what they already know in order to solve the learning tasks. Accordingly, supportive information provides mental models and cognitive schemata, that allow for multiple uses of the same general knowledge for performing different tasks. Procedural information is a prerequisite to the learning and performance of recurrent aspects of the learning tasks in each task class. It allows students to complete and learn routine aspects of learning tasks by specifying exactly how to solve the routine aspects of the tasks. It is presented just in time when learners need it. The procedural information may include hints (e.g. a summary of theory) or feedback relevant for the specific problem while working on the learning task. A lot of support is given for learning tasks early in a task class, but the support diminishes as learners acquire more expertise. Furthermore, part-task practice (e.g. drill- and practice exercises) supports the more complex whole task learning by providing additional exercises for selected recurrent constituent skills (van Merriënboer et al., 2003). Learners should be guided during the learning process by giving them summative and formative feedback (Elen, 2015). Accordingly, they should be able to perform, assess, and select tasks that fulfill their personal needs (van Merriënboer & Sluijsmans, 2009).

Nevertheless, providing an online learning environment with a well-considered instructional design does not ensure its effectiveness (Jiang et al., 2009). The effectiveness of the online learning environment is strongly related to students' capacity to properly use the different components (Lust et al., 2013). An important indicator of students' appropriate use is prior knowledge (van Merriënboer & Sluijsmans, 2009). In addition, the effects of students' motivational aspects on their willingness to use the components should not be ignored (Rienties et al., 2009). In the next section the influence of prior knowledge and motivation on use will be discussed based on prior research findings.

Students' Cognitive and Motivational Characteristics

A 4C/ID-based online learning environment is claimed to stimulate self-directed and deep learning by providing different components at the student's disposal, containing learner support on the one hand and by offering a lot of learner control on the other (Moos & Azevedo, 2009; van Merriënboer & Sluijsmans, 2009). Moreover, by giving students control of the use of the different components, adaptive learning based on their learning needs, should be possible. Despite this claim, we barely know if students actually benefit from these learning opportunities. Providing learning opportunities within an online learning environment is not sufficient to achieve better learning outcomes. A possible learner characteristic that could influence differences in use of the different components is students' prior knowledge (van Merriënboer & Sluisjmans, 2009). Based on CLT, students with low prior knowledge cannot immediately be confronted with highly difficult learning tasks. Accordingly, students' cognitive load can be reduced by consulting support and guidance (Jiang et al., 2009; van Merriënboer & Sweller, 2005). This would imply that students who perceive the task as more difficult use the components more or differently. Nevertheless, selecting various support or making additional exercises can be very challenging for students with low prior knowledge as they are more likely to perceive high cognitive load (Clarebout et al., 2010; van Merriënboer & Sluijsmans, 2009; van Merriënboer & Sweller, 2005). As a low level of prior knowledge increases cognitive load when faced with difficult learning tasks, those students might encounter problems to self-direct their learning. Accordingly, based on CLT, students with higher prior knowledge are more capable to self-direct their learning compared to students with lower prior knowledge (Moos & Azevedo, 2009; van Merriënboer & Sluijsmans, 2009). This would imply that students' prior knowledge can influence differences in the use of the four components. However, there is no solid empirical basis for this theoretical claim.

Several studies investigated the influence of students' prior knowledge on differences in use. Van Seters, Ossevoort, Tramper, and Goedhaert (2011) used online learning materials to demonstrate how university students work differently based on their prior knowledge. They investigated the learning path students followed when working with adaptive online learning material. The learning path was determined by the average step size chosen, the average number of tries, the total number of exercises, and the time needed to finish. They found that prior knowledge did not have an effect on students' learning path. Furthermore, Taub et al. (2014) studied the specific use of an online learning environment. Participants were 112 undergraduate students. Results revealed that all students visited a similar number of relevant pages regardless of their level of prior knowledge. Additionally, Jiang et al. (2009) conducted a study in which they measured variety in non-embedded tool use in an online learning environment (e.g., checklist tool, information list, calculator etc.). Tool use was measured by the frequency of tool use and proportional time spent on tools. They found no influence of prior knowledge on the quantitative aspects of tool use. Notwithstanding, research has shown that prior knowledge plays a primary role in learning achievement. Song, Kalett, and Plass (2016) studied the direct and indirect effects of university students' prior knowledge on learning performance in online learning environments. SEM revealed that university students' prior knowledge directly positively affected their learning outcomes. These aforementioned studies seem to confirm that students do not grasp learning opportunities based on their level of prior knowledge, but they do indicate the important role of students' prior domain knowledge on students' learning gain.

Additionally, motivational characteristics can have an important influence on students' learning behavior in online learning settings (Chen & Jang, 2010; Rienties et al., 2009). According to Expectancy-Value theory, self-efficacy and task value are two key components for understanding students' specific use and academic outcomes (Liem et al., 2008). Self-efficacy is defined as a learners' ability to execute the required behavior necessary for success (Greene & Azevedo, 2007). There is evidence that self-efficacious students participate more readily, work harder and persist longer when they encounter difficulties than those who are uncertain about their capacities (Zimmerman, 2000). Task value essentially refers to the reason for doing a task. More specifically, students with high task value pursue the enjoyment of learning and understanding new things (Joo, Lim & Kim, 2013).

Martens, Gulikers, and Bastiaens (2004) investigated the impact of task value on the use of online learning environments. The participants were 33 higher education students. Results showed that students with high task value did not do more, but did other things than students with low task value. Analysis of log files showed that students with high levels of task value showed proportionally more explorative study behavior. Explorative pages were defined as pages that students were not explicitly directed to by the external source. Studies also indicate relationships between self-efficacy, task value and performance. Bong (2001) conducted a path analysis to investigate the relationships between task-value, self-efficacy and performance (i.e. students' mid-term scores) in an online learning context. Participants were 168 undergraduate university students. Results showed strong links of self-efficacy with performance, whereas task value was linked to course enrollment intentions (Bong, 2001). Similarly, Joo et al. (2013) investigated 897 learners in an online university course to unravel relationships between self-efficacy, task value and performance. Using SEM they found significant positive relationships between both task value, and self-efficacy on course performance (i.e. the final grade on the course: midterm exam, attendance, and final exam). However, Song et al. (2016) conducted a SEM-analysis to examine the direct effects of task value and self-efficacy on students' learning outcomes, measured by a knowledge post-test. Participants were 368 university students. Results revealed no direct influence of self-efficacy and task value on students' learning outcomes. Major difference with the study of Joo et al. (2013) is that important additional predictors of students' learning outcomes were included in the research model such as prior knowledge and self-regulation. In sum, based on these aforementioned theoretical and empirical claims, we hypothesize that prior knowledge, self-efficacy, and task value influence differences in use of the four components and that students' cognitive and motivational characteristics can influence students' learning gain. Therefore, we formulate the following first research question (RQ):

- ***RQ1:*** How do students' cognitive (i.e. prior knowledge) and motivational (i.e. self-efficacy and task value) characteristics influence the use of the four components of a 4C/ID-based online learning environment?

Differences in the use of online learning components can influence students' performance. Lust et al. (2012) conducted a literature study which provided empirical evidence for the beneficial influence of differences in tool use (e.g. information, processing, and scaffold tools), on students' performance. Therefore, in this study activity data of the four different components is studied separately in order to gain more insight into how students' differences in the use of the four components contribute to performance, controlling for students' cognitive and motivational characteristics. Subsequently, the following second research question is formulated:

- ***RQ2:*** Do differences in the use of the four components of a 4C/ID-based online learning environment influence students' learning gain, taking into account students' cognitive and motivational characteristics?

Based on these research questions a theoretical research model is proposed as shown in *Figure 1.* containing all variables in order to elucidate the relationships among these variables.

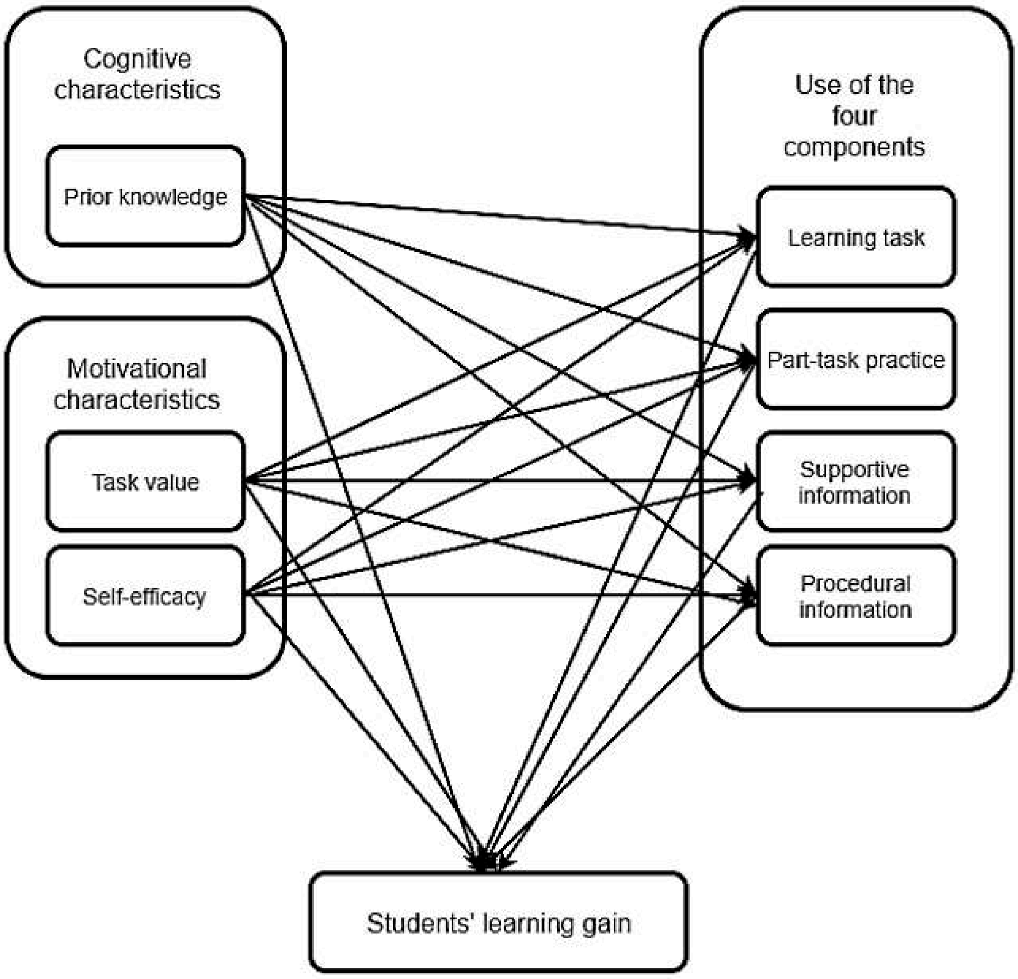


Figure 4.1: Research model

4.3. Method

The online learning environment

The online learning environment in the present study focuses on French as a foreign language. In Flemish education students generally start learning French at the age of 10 in Grade 5. French is always the first second language that is introduced into the curriculum and remains a compulsory component throughout primary and secondary education. Accordingly, the level of difficulty was aligned with the level that students in the Flemish part of Belgium are expected to reach at the end of secondary school. The main topic is 'traveling in France'. The learning environment covers four task classes. Each task class focuses on a theme (e.g. ordering your food in a restaurant). The online learning environment takes about 1 hour and 15 minutes to complete. The learning environment contained multiple media: videos, recorded audio and articles from different websites and is designed within the Moodle open-source learning platform. The instructional design of the online learning environment was developed in line with the guidelines of the 4C/ID-model. The learning tasks were based on authentic situations and sequenced in a simple-to-complex order. Each learning task started with an introduction where a worked example was given combined with clearly defined course objectives. Students received automatically generated feedback based on their scores (i.e. information about their achievements and guidelines). When students had an insufficient score they were advised to consult additional support and/or consult additional tasks in order to pass the learning task. When students finished their learning tasks, they had to hand in an assignment (e.g. writing down a conversation). The assignments were corrected by the instructor. Additional support and tasks were provided by the other three components of the 4C/ID-model, namely, supportive information (e.g. grammar explained by theory), procedural information (e.g. grammar explained by using keywords), and part-task practice (e.g. drill-and practice exercise of specific grammar). The supportive information, procedural information and part-task practice were non-embedded. More specifically, they were at the disposal of the students but the students had to decide whether or not to use them, and when they wanted to use them. Students had the opportunity to click on links to watch procedural or supportive information or to consult additional part-task practice (automatically rated by Moodle). In conclusion, consulting supportive and procedural information, making additional part-task practice, or retrying a learning task was the students' responsibility. Learning tasks were partly non-embedded since students were free to make as many learning tasks (e.g. several attempts) as they wanted. Nevertheless, as aforementioned, they were also partly embedded (i.e. less optional) since students were strongly advised to complete the learning tasks during the first administration session and since they were clustered and sequenced in a predefined order within a task class.

Participants

The study took place in the Flemish part of Belgium, at a Flemish university. The participants were 161 first-year Psychology and Educational Science students. The majority of the students were female (91%).

The average participant was 20 years old (*SD* = 2.92). Two students were perfectly bilingual (i.e. French and Dutch speakers). One student was German and had never learned French. Participation in research is part of the students' training program, but French was not a part of their training program. Before answering the questions, informed consent was obtained from all individual participants included in the study.

Study design

The design of the study consisted of two administration sessions. The first administration session started with an introduction of the online learning environment and self-reported questionnaire on task value and self-efficacy. Task value and self-efficacy were measured after the introduction of the online learning environment to make sure students' had sufficient insight into the learning content. The students were asked to use the learning environment at home for two weeks. As the learning content was not a part of their training program, they received the instructions that consulting the four components was optional and that there was no strict trajectory on how to work in the online learning environment. Nevertheless, consulting the learning tasks was strongly recommended. This implies that a lot of learner control was given to the students.

Measurements

Prior knowledge and students' learning gain. To measure students' prior knowledge and students' learning gain of French a quantitative paper-and-pencil instrument constructed by Evens et al. (2017) was used as a pretest and posttest. The instrument consists of 60 items and focuses on knowledge (i.e. grammar and vocabulary) and skills (i.e. listening, writing a conversation). The level of difficulty of the test was B1 of the Common European Framework of Reference. In this study, students' learning gains are the results of the posttest, controlled for the pretest. The instruments' reliability was explored by calculating internal consistency, that is, Cronbach's α = .90 for the pretest and Cronbach's α = .89 for the posttest. Both results reveal a good internal consistency.

Self-efficacy and task value. Within this study the constructs self-efficacy and task value were retrieved from the motivated strategies for learning questionnaire (MSLQ) constructed by Pintrich and De Groot (1990). MSLQ consists of a motivation section and a learning section. For this study we used the constructs self-efficacy (e.g. "I expect to do well in this course"), and task value (e.g. "It is important for me to learn the course content"), of the motivation section. The questionnaire was a 5-item scale with a 7-point Likert-type response format having values ranging from strongly agree (7) to strongly disagree (1). The questions were translated into Dutch. Construct validity was checked by conducting a confirmatory factor analysis (CFA). The CFA model relates observed responses or 'indicators' to latent variables (i.e. measurement model). CFA indicated that the measurement model exhibited good validity. The standardized factor loadings from the latent variable constructs were all significant with standardized values ranging from .73 to .93. and an average variance explained (AVE) of .76 for self-efficacy and .62 for task value. Therefore, we can suggest that the two measurement models for each construct were measured well in the current data. Internal consistency was investigated by measuring Cronbach's Alpha. The Cronbach's Alpha for self-efficacy was α = . 94 and for task value α = .84, which indicates high reliability (Rosseel, 2012).

The use of the four components. Information of students' use of the four components was collected by tracking students' activity. Students' activity includes any kind of interaction (e.g. views, attempts, submitting quizzes etc.) with the online learning environment during two weeks, tracked for each component of the 4C/ID-model separately. All data were anonymized through means of the use of random codes to safeguard the identities of the students. User activity was chosen instead of time spent because it gave more accurate information about the use of supportive and procedural information.

Analysis

Firstly, descriptive analyses such as mean, standard deviation and correlation analysis were conducted for the different variables (i.e. students characteristics and the four components of the 4C/ID model). Secondly, in order to find answers on the first research aim the effect of students' motivational and cognitive characteristics (i.e. latent variables), on the use of the four components (i.e. manifest) variables, was investigated by conducting SEM in R using the Lavaan package 3.4.0 (Rosseel, 2012). Figure 1 was specified as the statistical model using the latent variables as shown in Fig. 2 . SEM is a statistical approach to test hypotheses about the relationships among observed/manifest (i.e. rectangles) and latent variables (i.e. ovals). As shown in Figure 2, self-efficacy and task value are the latent constructs. In the current study, in order to answer the first research question, students' cognitive (i.e. prior knowledge) and motivational (i.e. task value and self-efficacy) characteristics are the independent variables and the four components (i.e. learning task, part-task practice, supportive and procedural information) are the dependent variables. Additionally, to give answer to the second research question, the use of the four components, and students' cognitive and motivational characteristics are the independent variables, and students' learning gain the main dependent variable.

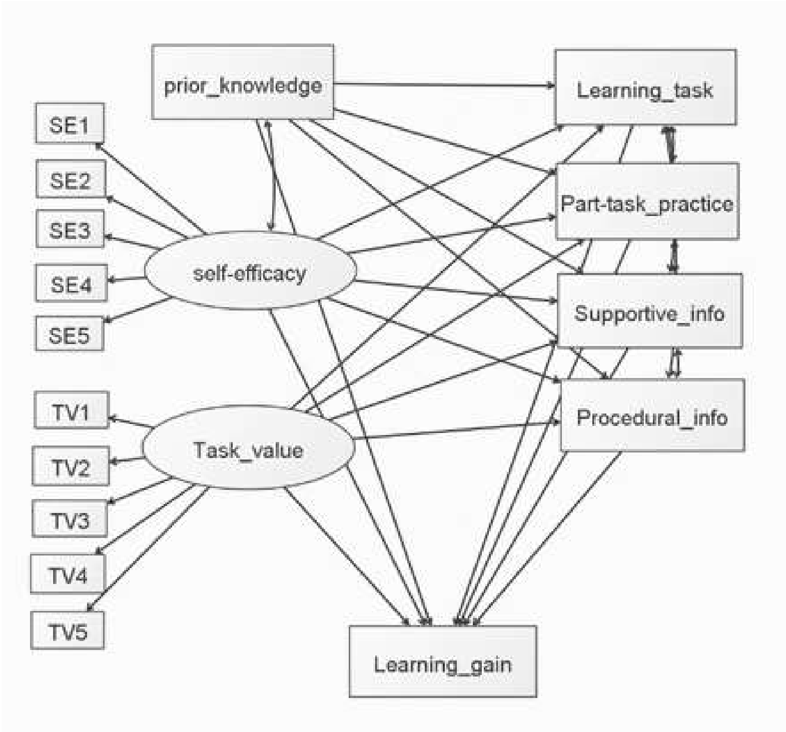


Figure 4.2: Statistical model

4.4. Results

Preliminary analysis

Results of the descriptive statistics of students' pretest and posttest can be found in Table 1. The average results on students' pretest was 52.32%. The average score on the posttest was 64.53%. This indicates that using the online learning environment improved their performance. Nearly all students consulted the learning tasks (N = 158). Not all students consulted supportive information (N = 125), procedural information (N = 140) and/or consulted part-task practice (N = 72). The average time spent on using the online learning environment was 66 minutes (SD = 27.34, *min.* = 10.44 minutes, *max.* = 151.43 minutes).

Table 4.1.

Descriptive statistics of the manifest variables

Variable	*N*	*Min*	*Max*	*Mean*	*SD*
Pretest	151*	7.8	92.19	52.32	17
Posttest	152*	20.97	98.44	64.53	14.81
Learning task	158	15	180	80.60	25.90
Part-task Practice	161	0	85	10.50	19.85
Supportive info	161	0	64	9.04	10.94
Procedural info	161	0	36	9.30	8.52

**not all students were present during the pretest/posttest. Results of the two bilingual students and the German student were removed as outliers.*

Table 2 gives an overview of the correlations among the variables. RQ1 investigates the influence of students' cognitive (i.e. prior knowledge) and motivational (i.e. task value and self-efficacy) characteristics on the use of the different components (i.e. use of the learning tasks, part-task practice, supportive and procedural information). There is a negative significant correlation between prior knowledge and part-task practice. Additionally, there is a negative significant correlation between prior knowledge and supportive information. There is a positive significant relationship between task value and learning tasks. Additionally there is a positive significant relationship between task value and supportive information. There are no significant relations between self-efficacy and use of the four components. RQ2 investigates the influence of the use of the different components on students' learning gain, taken into account students' cognitive and motivational characteristics. There is a positive significant correlation between learning task and posttest. Additionally, there is a positive significant correlation between procedural information and posttest. There is a negative significant relationship between posttest and part-task-practice. Results also reveal that there is a positive significant correlation between pretest and posttest.

Table 4.2.

Correlations among the variables

	1	2	3	4	5	6	7	8
1. Pretest	1							
2. Self-efficacy	.46**	1						
3. Task value	.06	.28**	1					
4. Learning task	-.12	.06	.22**	1				
5. Part-task practice	-.22**	-.00	.04	.31**	1			
6. Supportive Information	-.21*	-.12	.18*	.38**	.58**	1		
7. Procedural Information	-.08	-.07	.08	.22*	.09	.25**	1	
8. Posttest	.86**	.38**	.11	.02	-.19*	-.15	.01	1

***correlation is significant at the .01 level; *correlation is significant at the .05 level*

Structural Equation Model

SEM was conducted in order to investigate the relationships between students ' cognitive and motivational characteristics, the four components of the 4C/ID-model and students' learning gain (i.e. RQ1 and RQ2). For the missing values a two-stage approach was applied. This approach obtains a saturated maximum likelihood (ML) estimate of the population covariance matrix and then uses this estimate in the complete data ML fitting function to obtain parameter estimates (Savelei & Bentler, 2009). Lavaan converged normally after 59 iterations. The hypothesized model, provided an adequate fit to the given data. The χ^2-test indicates the difference between observed and expected covariance matrixes and should be non-significant. However, χ^2-test is highly dependent on sample size and therefore normed χ^2-test is often considered, this is, χ^2-test divided by the degrees of freedom (*df*). Values smaller than 2.0 are considered to indicate acceptable fit (Rosseel, 2012). In addition to χ^2 statistics, the root mean squared residual (SRMR), the root mean squared error of approximation (RMSEA), comparative fit index (CFI) and the Tucker-Lewis Index (TLI) were examined. Table 3 summarizes the overall goodness-of fit measures of the model. SRMR is the difference between the observed variance and the predicted variance. A value less than .06 is considered a good fit. RMSEA is related to residuals in the model, by adjusting for the complexity of the model and the size of the sample. A marginal value for acceptance is < .08. CFI is the discrepancy function adjusted for the sample size. A value of CFI and TLI between > .95 indicates good fit. Assessing all measures and considering the above statements, the original structural model was accepted and considered adequate for further analysis (Hu & Bentler, 1999; Kline, 2013; Rosseel, 2012).

Table 4.3.

Model fit

Fit measures	Values	Recommended value
Chi-square (χ^2)	155.56 (*df* = .99, *p* = .00)	non-significant
Normed Chi-square	1.6	X^2/df <.02
SRMR	.05	< .06
RMSA	.06	< .08
CFI	.97	> .95
TLI	.95	> .95

Using SEM analysis we firstly investigated the influence of students' characteristics on the four components of the 4C/ID-model (i.e. RQ1). Fig. 3 gives an overview of the standardized path coefficients. The solid lines show the significant relationships. Significant relationships between students' cognitive and motivational characteristics and use of the four components were found. More specifically, a significant negative influence was found of students' prior knowledge on the use of part-task practice (β = -.21, p <.05). No significant relationships were found between students' self-efficacy and the use of the four components of the 4C/ID-model. Task value positively influences the use of learning task (β = .21, p <.05) and supportive information (β = .22, p <.05). The variance explained for the learning task is respectively (R^2 =.07), for part-task practice (R^2 =.03), for supportive information (R^2 =.08) and for procedural information (R^2 =.02).

RQ2 investigated the influence of the use of the four components on students' learning gain, taking into account students' characteristics. Significant relationships were found between the use of different components and students' learning gain. Respectively, a significant influence of the use of learning tasks (β = .12, p <.01) and procedural information (β = .08, p <.05) on students' learning gain was found. A positive significant relationships between students' prior knowledge on students' learning gain was found (β = .91, p <.001). No further relationships between students' motivational characteristics and their learning gain were found. The variance explained for students' learning gain was (R^2 = .79). In conclusion, students' use of the components of the 4C/ID model is negatively influenced by students' prior knowledge and positively by students' task value. Secondly, results indicate that when using a 4C/ID-based online learning environment, mainly using the learning tasks and procedural information contributes to students' learning gain. Additionally, students' learning gain is mainly influenced by their prior knowledge.

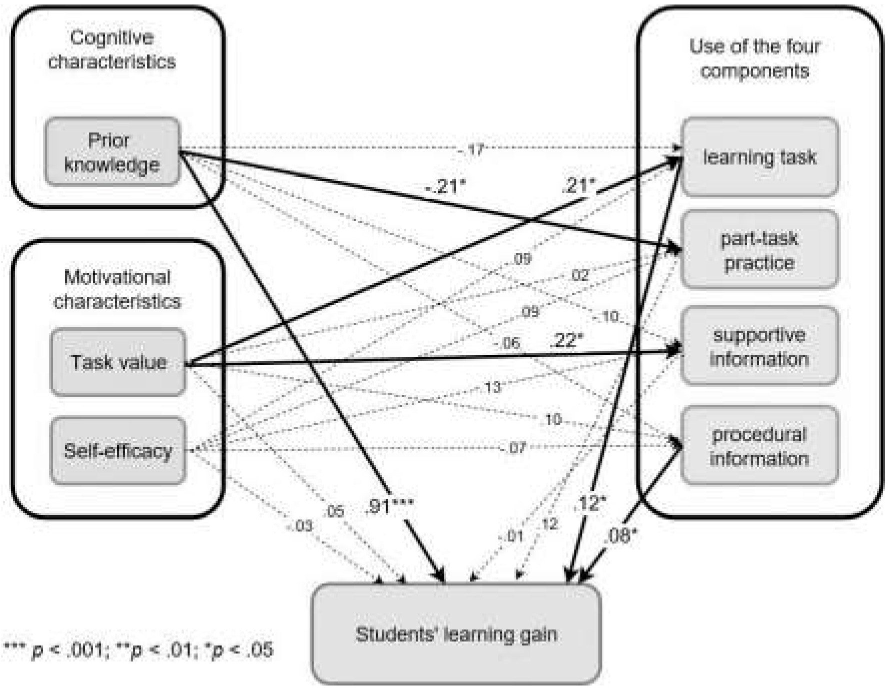


Figure 4.3.: Structural model with standardized path coefficients

4.5. Discussion

The current study strived to investigate (1) the influence of students' cognitive (i.e. prior knowledge), and motivational (i.e. task value and self-efficacy), characteristics on students' differences in use of the four components (i.e. RQ1) and (2) the influence of students' differences in use on students' learning gain, taking into account students' cognitive and motivational characteristics (i.e. RQ2). All variables were incorporated in a structural research model. Our results are based on data from a pretest (i.e. prior knowledge), self-reported questionnaires (i.e. task value and self-efficacy), platform log data from 161 students (i.e. activity of the four components of the 4C/ID-model) and a posttest, controlled for the pretest (i.e. students' learning gain). RQ1 investigates the influence of students' cognitive and motivational characteristics on differences in use of the four components. Results indicate that students' prior knowledge has a negative significant influence on the use of part-task practice. More specifically, results reveal that the lower students' prior knowledge was, the more students consulted additional part-task practice. Part-task practices contain additional exercises with more recurrent content in a drill and practice- format in order to prepare students to solve the learning tasks which contain both recurrent and non-recurrent content. Therefore, the current findings indicate that in general students seem to be aware that they are lacking routine knowledge to solve the learning tasks, and subsequently they self-direct their learning in order to achieve better results. Taking into account cognitive load theory, these results could imply that students did not experience high intrinsic cognitive load, as they were still able to self-direct their learning (van Merriënboer & Sweller, 2005). The current findings are in contrast to the study of Taub et al. (2014) which indicated that regardless of the sub-goals students were working on, there were no significant differences in students' use of the online learning environment (i.e. defined by the number of relevant pages visited) between lower and higher prior knowledge groups. Current findings are also different from the study of Jiang et al. (2009) who measured differences of use based on prior knowledge by looking at the frequency of tool use and proportional time spent on tools, but found no differences (Jiang et al., 2009; Song et al., 2016).

A possible explanation for the different findings could be found in the respective procedure of the studies. Concretely, in both studies students worked with the online learning environment under supervision or in a controlled setting. By controlling the setting students might feel the pressure to complete different tasks in a given time which could result in following a more traditional linear path. Moreover, important self-regulation skills that are typical for online learning are not taken into account when the setting is controlled, such as time management (i.e. the ability to effectively manage learning) and environment structuring (i.e. being able to structure your own learning environment; Barnard-Brak, Paton & Lan, 2010; Whipp & Chiarelli, 2004). These self-regulation skills could have an influence on differences in use between different prior knowledge groups. Furthermore, looking at the study of Taub et al. (2014) defining differences in use based on relevant pages visited might have been too narrow. In the study of Taub et al. (2014), multichannel data was collected including log-files, but also think-aloud protocols, electrodermal activity (EDA), facial expressions, and eye-tracking to measure metacognitive and cognitive self-regulation. Based on multichannel data results did reveal that metacognitive self-regulation processes differed between lower and higher prior knowledge groups which indicates that prior knowledge groups do differ in how they use the online learning environment and/or self-direct their learning, but that the differences were not visible based on the relevant pages visited. Not only prior knowledge but also motivational characteristics seem to influence differences in students' use of the different components. Results indicate differences in activity based on students' task value. Prior research on expectancy-value theory has indeed shown that after controlling for prior knowledge, task value can predict differences in academic decisions (Greene & Azevedo, 2007). Students' task value seems to have a positive influence on the activity of the learning tasks and supportive information. As supportive information provided broader background information this could imply that students with higher task value put more effort in solving the learning tasks qualitatively and/or are more eager to learn. The findings correspond with the study of Martens et al. (2004) which also analyzed log files and found that students with high intrinsic motivation showed more explorative study behavior. Explorative study behavior in their study was calculated by dividing the number of explorative pages a student had visited by the total number of visited pages (Joo et al., 2013). Studying the influence of self-efficacy on the activity of the four components, no link between self-efficacy, and the use of the four components was found. As correlations indicated that self-efficacy and prior knowledge were highly correlated. This indicates that in general students from this study could properly assess themselves. Moreover, this indicates that they did not feel the need to self-direct their learning in order to reach a certain quality standard (Greene & Azevedo, 2007).

RQ2 investigated the impact of differences in use of the four components on students' learning gain, taken into account students' characteristics. Results revealed a significant influence of students' use of the online learning environment on students' learning gain. Results indicate that activity of the learning tasks, and procedural information contributed to students' learning gain. These results indicate that the more learning tasks and procedural information were consulted, the higher students' learning gain. Procedural support and guidance can prevent learners from paying attention to irrelevant task aspects, and is therefore argued to reduce extraneous cognitive load, which on its turn can improve students' task performance (van Merriënboer & Sluijsmans, 2009; van Merriënboer et al., 2003). This could be the reason why procedural information contributes to performance within this study. Activity of the supportive information and part-task practice did not contribute to course performance. This could be due to the fact that the learning tasks were not complex enough. Consulting procedural information must have been sufficient for students to solve the learning tasks. Accordingly, students did not feel the need to consults these components. When investigating the influence of differences in use on students' learning gain, it is important to take students' characteristics into account. Results revealed a major significant influence of students' prior knowledge on students' learning gain. Correlations already indicated that pretest and posttest were highly correlated. This indicates that students' learning outcomes were mainly influenced by students' prior knowledge. This is in line with previous research. The study of Song et al. (2016) already revealed a direct effect of prior knowledge on performance (Taub et al., 2014). There was no effect of students' motivational characteristics on students' learning gain.

More specifically, no significant effect of students' self-efficacy on students' learning gain was found. These findings are in contrast with the study of Joo et al. (2013) but are in line with those of the study of Song et al. (2016). Song et al. (2016) did not find a significant relationship between self-efficacy and students' learning outcome. A possible explanation for these different findings is that the online learning environment used in their study was highly complex (Song et al., 2016; Zimmerman, 2000). Former studies have indicated that efficacious students can be less confident when they are challenged with challenging learning environments (Greene, Moos, Azevedo & Winters, 2008). Different outcomes can also be influenced by the design of the research model. In contrast to the current study and the study of Song et al. (2016), Joo et al. (2013) did not incorporate prior knowledge into the research model as a predictor of students' learning outcome. Furthermore, students' task value has no influence on students' learning gain. However, this finding is in contrast with the study of Joo et al. (2013). They have conducted SEM analysis with self-efficacy and task value as predictors of performance and found a direct influence of task value on students' learning outcome (Zimmerman, 2000). It should be noted that the students in the study of Joo et al. (2013) were enrolled in an online university, which possible had an influence on the level of students' motivation. Therefore, possible clarification for the different findings is that students' task value in the current study was generally too low to have an effect on performance since the online course was not a part of the students' study program.

Limitations and future research

A first important limitation of this study is that students' cognitive (i.e., prior knowledge) and motivational characteristics (i.e. task value and self-efficacy) explain little variance of the use of the different components. This implies that more predictors should be incorporated into the model to predict differences in use. In the current study a possible predictor of differences in use and students' learning gain is students' self-regulation skills. The influence of students' self-regulation skills is very important since online learning offers a lot of learner control and therefore gives the students a great degree of autonomy (Song et al., 2016; Tsai, 2013). Incorporating self-regulation skills in the research model could give more insight in why students use specific components during their learning process. A second important limitation is that little is known about the way in which the different components are used. By analyzing log-data more in detail, such as, looking at the sequences of use of the four components, more insight could be given on effective use. A possible example of effective use could be that when a student has an insufficient score, the student decides to consult supportive information.

Subsequently, this detailed information could give more insight into their self-directed learning. A third limitation is that one component can contain a lot of different learning activities (e.g. quiz, reading and/or, listening activity). More detailed studies are needed to measure the influence of the various learning activities. This could provide more insight in how to design tailor-made online learning environments (Rienties & Toetenel, 2016). Fourthly, the current study was conducted in an experimental setting. Accordingly, students mainly used the online learning environment to accomplish the learning tasks, but not to study the specific content. More research needs to be done in contexts in which courses are actually part of the training program in order to have more knowledge about how students actually study in a 4C/ID-based online learning environment. A fifth limitation is the cross-sectional design of this study. It would be interesting to conduct longitudinal research instead. This would provide insight into how students study within an online learning environment and how this influences students' learning and educational outcomes in the long term. Finally, future work should replicate and extend the current findings with other 4C/ID-based online learning environments and other target groups, to test generalizability. Former research has shown that students' use of online resources and their performance varied between courses and that this strongly depends on the instructional design (Jeong & Hmelo-Silver, 2010).

4.6. Conclusion

This study firstly provides more information about the influence of students' motivational and cognitive characteristics on the use of the four components in a 4C/ID-based learning environment. In general, results indicate that students' characteristics do influence differences in use of the four components when students receive a lot of learner control. Moreover, the 4C/ID-model seems to be an instructional design model that is adequate for students with different characteristics. It allows students to self-direct their learning by providing four components that can be consulted freely in a non-linear trajectory. Furthermore, the use of learning tasks and procedural information controlled for students' prior knowledge, seems to influence students' learning gain directly. This indicates the importance of combined use of learning tasks and procedural information. More insight into how students differ in use, based on their characteristics is an important step from an instructional design perspective. This could provide important suggestions for designing online learning environments that improve students' learning gain by allowing them to self-directed their learning to fulfill their personal needs.

Chapter Five

Combining physiological data and subjective measurements

Larmuseau, C., Cornelis, P. Lancieri, L., Desmet, P., & Depaepe, F. (2020). Multimodal learning analytics to investigate cognitive load during online problem solving (special issue: The Promise and Challenges of Multimodal Learning Analytics. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.

5. Combining physiological data and subjective measurements

5.1. Introduction

As society and work environments become more complex it is increasingly relevant that learning environments mirror this complexity of the real world (Jonassen, 2000; Kirschner, Ayres & Chandler, 2011; Merrill, 2009; van Merriënboer et al., 2003). Nevertheless, a risk of complex learning environments is that the cognitive load imposed by the complex learning tasks is often excessive (van Merriënboer & Sluijsmans, 2009). This phenomenon can be explained by Cognitive Load Theory (CLT) introduced by Sweller (1994). CLT uses current knowledge about human cognitive architecture as a baseline to develop the instructional design for complex learning environments (Martin, 2014). CLT distinguishes three types of cognitive load, intrinsic, extraneous, and germane load (Brunken, Plass & Leutner, 2003; Paas, Tuovinen, Tabbers & Van Gerven, 2010; Sweller, 2010). The level of intrinsic load is assumed to be determined by the complexity of the task or learning material and cannot be directly altered by the instructional designer. Extraneous load is mainly imposed by instructional procedures that are suboptimal, whereas germane load refers to the learners' working memory resources available to deal with the complexity of the task or learning material (Sweller, 2010). Both extraneous and germane load can be facilitated by the instructional designer. An instructional designer should find a balance between keeping the matter sufficiently challenging but still within the cognitive capacities of the learner. Exceeding learners' cognitive capacities can induce cognitive overload which could hamper learning. Specifically, this means that when the content is very complex due to high element interactivity (i.e., the number of interrelations between knowledge, procedures, formulas etc.) which affects intrinsic load, instructional designers should keep the extraneous load to a minimum (e.g., by providing clear instructions, provide embedded support) and subsequently foster germane load (Kirschner et al., 2011; Sweller, 2010).

In order to align the instructional design with students' cognitive abilities, we should be able to measure cognitive load during complex learning. Former studies investigated cognitive load by using subjective measurements such as self-reported questionnaires (Boekaerts, 2017; Zheng & Cook, 2012). Those self-reported questionnaires have some important disadvantages (e.g. subjective measures, assumption of constant workload capacity, DeLeeuw & Mayer, 2008; Raaijmakers, Baars, Schaap, Paas & van Gog, 2017). As a result, more researchers show interest in using objective, real-time measures. Physiological measures provide objective data and can be unobtrusively collected while dealing with a task or learning material. Moreover, physiological data might indicate changes in cognitive functioning throughout the process of solving a task (Boekaerts, 2017). Former studies already indicated that electrodermal activity (EDA) and skin temperature (ST) can be linked to different levels of task complexity (Haapalainen, Kim, Forlizzi & Dey, 2010; Nourbakhs, Wang, Chen & Calvo, 2012; Shi, Ruiz, Taib, Choi & Chen, 2007).

Nevertheless, it is unclear whether these physiological measures are related to self-reported intrinsic load, extraneous load, germane load and the overall mental effort during complex problem solving (Leppink, Paas, Van der Vleuten, Van Gog & Van Merriënboer, 2013). Therefore, in the current study, a high and low complex task was developed relating to the learning and teaching of geometry. The complexity of the task was manipulated by increasing the element interactivity for the high complex task (Sweller, 2010). In both tasks the same amount of support was provided. Data was retrieved using self-reported questionnaires to measure students' experienced intrinsic load, extraneous load, germane load and mental effort. This distinction between the different types and mental effort was made because the different types of cognitive load concerns mental load induced by task complexity and instructional design, whereas mental effort invested covers the overall amount of cognitive processing for a particular task (Paas et al., 2003). The subjective measures were combined with physiological data through wrist-worn wearables containing both EDA and ST.

The purpose of this study was threefold. First, we investigated differences in the experienced cognitive load and the physiological data while solving a high and low complex task. Secondly, we examined whether individual differences of subjective measurements are related to individual differences of physiological data for the high and low complex task. Finally, we described whether peaks (i.e. EDA) and/or drops (i.e. ST) of physiological data are related to specific events (e.g. consultation of support) that took place during the problem solving process.

5.2. Theoretical background

Cognitive Load Theory

CLT is concerned with the instructional implication of the interaction between the complexity and instructional design of the learning material and human cognitive architecture (Sweller, 2010). Basically, the human cognitive architecture consists of an effectively unlimited long-term memory, which interacts with a working memory that has limited processing capacity (Kirschner et al., 2011; Sweller, 1994). Long-term memory contains cognitive schemata that are used to store and organize knowledge. Learning occurs when information is successfully processed in working memory and when new schemas are created or incorporated into consisting schemas in the long-term memory. As the processing capacity of the working memory is so limited, overcoming individual working memory limitations by instructional manipulations has been the main focus of CLT (Sweller, van Merriënboer & Paas, 1998). Cognitive load can be defined as a multidimensional construct representing the load that performing a particular task, imposes on the learners' cognitive system (Paas et al., 2010). CLT claims that the cognitive load that learners experience can be intrinsic, extraneous or germane (Sweller, 2010). The level of intrinsic load for a particular task is assumed to be determined by the inherent difficulty of a certain topic and the level of element interactivity of the learning material in relation with student's prior knowledge. The more elements that interact, the more intrinsic processing is required for coordinating and integrating the material and the higher the working memory load (De Leeuw & Mayer, 2008; Sweller, 2010).

Working memory load is not only imposed by the intrinsic complexity of the material that needs to be learned, it can also be imposed by the instructional design. For instance, unclear instructional procedures can impose extraneous load. Extraneous processing means that the learner engages in cognitive processing that does not support the learning objective (De Leeuw & Mayer, 2008; Glogger-Frey, Gaus & Renkl, 2017; van Merriënboer & Sluijsmans, 2009; Sweller, 2010). Instructional design techniques that reduce extraneous load (e.g. fading support) should ensure that students devote less attention to irrelevant aspects of the task. Subsequently, more cognitive capacity can be allocated to the actual learning objective (Ciernak, Scheiter & Gerjets, 2009; Mayer & Moreno, 2010; Sweller et al. 2011). Meanwhile, intrinsic and extraneous load depend on the characteristics of the learning tasks or the instructional design, germane load is more concerned with the cognitive characteristics of the learner. More specifically, it refers to the available working memory resources to engage in knowledge elaboration processes and argumentation (Sweller, 2010). Accordingly, in order to optimize learning, learning tasks should be aligned with the learner's cognitive capabilities (Schmeck, Opfermann, van Gog, Paas & Leutner, 2015; Sweller, 2010). Measuring cognitive load during complex learning should provide more insight into how to align instructional design with students' cognitive capabilities.

Subjective measurements of cognitive load

Self-reports for measuring cognitive load are subjective measurements consisting of unidimensional and multidimensional scales. Unidimensional subjective rating scales have been used intensively in research and have been identified as reliable and valid estimators of cognitive load (Boekaerts, 2017; Chang & Yang, 2010; Leppink et al., 2013; Paas, 2003). The Paas's nine-point mental effort rating scale has been most frequently used in cognitive load research (Chen et al., 2016; Paas, 1992). Paas's nine-point mental effort rating scale requires learners to rate their mental effort immediately after completing a task (Paas, 1992). Mental effort is the aspect of cognitive load that refers to the cognitive capacity that is allocated to accommodate the demands imposed by a task (Paas et al., 2005). According to Paas, learners can introspect the amount of mental effort invested during a learning task. Subsequently, Paas claims that the learner's assessment can be used as an index of overall cognitive load (Chen et al., 2016). Nevertheless, this unidimensional scale gives little insight into the influence of the complexity of the task and the influence of the instructional design on cognitive load (Boekaerts, 2017; De Bruin & van Merriënboer, 2017; Klepsch, Schmitz & Seufert, 2017; Leppink et al., 2013). Accordingly, Leppink et al. (2013) and Klepsch et al. (2017), developed a subjective cognitive load scale in which they used multiple items for each type of cognitive load in order to get more specific information about intrinsic load, extraneous load and germane load. Despite the frequent use of self-reported scales to assess cognitive load, some critiques have been raised. Firstly, subjective measurements are based on the assumption that students are able to introspect on their cognitive processes and accordingly are able to self-report on their experienced cognitive load (Boekaerts, 2017; Schmeck et al., 2015). Secondly, as subjective scales are often administered after the learning task, subjective scales do not capture variations in load over time. Taking into account these limitations, it might be more interesting to combine subjective measurements with real-time objective cognitive load information (Boekaerts, 2017; Chen et al., 2016; Zheng & Cook, 2012).

Physiological measures of cognitive load

The physiological approach for cognitive load measurement is based on the assumption that any change in human cognitive functioning is reflected in the human physiology. Subsequently, in contrast to subjective measurements, physiological measures are continuous and measured at a high frequency (e.g., every second) and with high precision (Chen et al., 2016). Given the close relationship between cognitive load and neural systems, human neurophysiological signals are seen as promising avenues to measure cognitive load (Boekaerts, 2017; Chen et al., 2016). Former research has investigated the relationship between learners' cognitive load and their physiological behavior. The physiological measures that have been used to investigate cognitive load are among others heart rate by electrocardiography (ECG), brain activity by electroencephalography (EEG), eye activity (e.g. blink rate, pupillary dilation), EDA, heat flux and ST (Antonenko, Paas, Grabner & van Gog, 2010; Haapalainen et al. 2010; Scharinger, Soutschek, Schubert & Gerjets, 2015; Smets et al., 2018; Zagermann, Pfeil & Reiterer, 2016). Although a lot of physiological data, such as the brain and eye activity, has been proven to be highly effective for measuring cognitive load, these types of physiological data often require expensive sophisticated equipment that is highly obtrusive in measuring cognitive activities, especially in ecologically valid contexts (Chen et al., 2016; Scharinger et al., 2015).

Possible solutions to collect physiological data in an unobtrusive way is by means of wrist-worn wearables. These wearables can easily capture different physiological data such as EDA and ST and are less expensive compared to more sophisticated measures of physiological data (Chen et al., 2016). EDA involves measuring the electrical conductance of the skin through sensors attached to the wrist. Skin conductivity varies with changes in skin moisture level (i.e. sweating) and can reveal changes in the sympathetic nervous system (SNS). The slowly changing part of the EDA signal is called the skin conductance level (SCL) and is a measure of psychophysiological activation. SCL can vary substantially between and within individuals. A fast change in the EDA signal (i.e. a peak) occurs in reaction to a single stimulus (also referred to as galvanic skin response, GSR; Braithwaite, Watson, Jones & Rowe, 2013). Research has linked GSR variation to stress and SNS arousal. As a person becomes more or less stressed, the GSR increases or decreases respectively (Hoogerheide, Renkl, Logan, Paas & van Gog, 2019; Liapis, Katsanos, Sotiropoulos, Xenos & Karousos, 2015, Smets et al., 2018). Additionally, research has also linked GSR readings to cognitive activity, claiming GSR responses increase when more cognitive load is experienced (Ikehara & Crosby, 2005; Nourbakhs et al, 2012; Setz et al., 2010; Shi et al., 2007, Yousoof & Sapiyan, 2013). The study of Nourbakhs, Wang, Chen, and Calvo (2015) captured GSR data from different reading and arithmetic tasks. The arithmetic tasks contained four difficulty levels and 13 participants whereas the reading task contained three difficulty levels and contained 16 participants. Results of ANOVA indicated that both mean GSR and accumulated GSR yielded significantly different results throughout different task difficulty levels. Shi et al. (2007) investigated 11 subjects when dealing with four tasks divided in four distinct levels of cognitive load. Results revealed insignificant differences across the interactive models for mean GSR, but significant differences when using accumulated GSR. Yousoof and Sapiyan (2013) investigated whether cognitive load could be detected by mean EDA. In this experiment, 7 subjects had to solve three different programming tasks that were different in terms of complexity. Yousoof and Sapiyan found no conclusive results for mean GSR, indicating that the variation among the subjects was very different during one task.

In addition to EDA, ST can also reflect changes in SNS. Research claims that acute stress triggers peripheral vasoconstriction, causing a rapid, short-term drop in skin temperature. Moreover, stress can also cause a more delayed skin warming, providing two opportunities to quantify stress (Herborn et al., 2015; Karthikeyan, Murugappan & Yaacob, 2012; Smets et al., 2018; Vinkers, et al., 2013). Little research has used ST to assess cognitive load. Nevertheless, the study of Haapalainen et al. (2010) investigated the cognitive load of 20 subjects through GSR and heat flux data (i.e. rate of heat transfer). The subjects had to solve six elementary cognitive tasks that differed in difficulty. Afterwards, Haapalainen et al. (2010) evaluated the performance of each of the features in assessing cognitive load using personalized machine learning techniques (i.e. Naïve Bayes Classifier). Results indicated that they did not obtain satisfactory results for GSR. By contrast, they did find that across all participants heat flux was shown to be an indicator of differences in cognitive load. The findings of former studies indicate that EDA and ST can indicate differences in cognitive load, but none of these studies related physiological data with self-reported cognitive load.

Research aims

To conclude, physiological measures have some important advantages when compared to subjective measurements. These measures are more objective (i.e. not dependent on students' perceptions), multidimensional (i.e. different physiological measures are sensitive to different cognitive processes), unobtrusive (i.e. no additional requirements), implicit (i.e. collect data while students are working on their tasks) and continuous (i.e. provide information of cognitive processes during learning). Nevertheless, it can be difficult to interpret physiological data. Therefore, it would be interesting to investigate whether there is a relationship between subjective measurements of cognitive load and physiological data. The following research questions are formulated:

- ***RQ1:*** Does the manipulation of the level of complexity of a task, based on element interactivity, result in differences in perceived cognitive load and mental effort when controlled for prior knowledge?
- ***RQ2***: Does the manipulation of the level of complexity of a task, based on element interactivity, result in differences in physiological data, when controlled for prior knowledge?
- ***RQ3***: Is there a relationship between individual differences in self-reported data and individual differences of physiological data for a high and low complex task?
- ***RQ4***: Is there a relationship between the physiological data of one learner and his/her interactive behavior during the problem-solving process?

5.3. Method

Participants and study design

Participants were 15 future primary school teachers of which ten were female and five males (age between 18-24). All participants were first-year bachelor students (i.e. second semester). The study was highly ecologically valid as the study was orchestrated by the students' lecturer of the teaching mathematics course unit. Moreover, the intervention was integrated into the students' study program (i.e. primary school teacher training). The intervention consisted of a within-subject design and was conducted online in the Moodle learning management system (LMS). The intervention took place in the auditorium of their faculty where students could solve the tasks individually on their computer among their fellow students. This session was supervised by their lecturer and a researcher. Students first received an online questionnaire of which the timeframe (+/- five min.) to complete the first questionnaire was used as an adaption period in order to stabilize the wearable signals (i.e. baseline measurements). Next, all students had to solve a highly complex and a low complex task on preparing a lesson in geometry as shown in Figure 1. In order to control for order effects, (a) half of the subjects were exposed to the highly complex task during the first session and the low complex task during the second session, whereas for (b) the other half, the sequence was vice versa. More specifically, eight students started with the highly complex task and seven students started with the low complex task.

High and low complex task

The high and low complex tasks were developed in Moodle LMS. The scope of both tasks was designing a lesson preparation on the circumference of a circle for primary school children. This subject matter was not yet covered in previous lessons. Both tasks contained six elements where both aspects of PCK (i.e. inductive teaching strategy, choose teaching materials to support your lesson, aligning the topic of the lesson with the Flemish curriculum and integration of differentiation in your lesson in the classroom) and CK (i.e. formula of the circumference of the circle) was addressed. The complexity of the highly complex task was manipulated based on element interactivity (Sweller, 2010). In the highly complex task, students had to coordinate and integrate six elements consisting of CK and PCK in order to write a course preparation about the circumference of the circle, whereas the low complex task consisted of six questions where each element was addressed separately (see Figure 1). During both problems, the same support consisting of procedural and supportive information was provided. An example of procedural information can also be found in Figure 1 in the second box. Procedural information is provided just-in-time and concise. Supportive information is much more extensive and is comparable to the background theory. Both procedural and supportive information can be consulted by clicking on the words in italics.

High complex task: Write out (briefly) a lesson preparation around π and the circumference of the circle (15 lines)

Targets of this lesson:
- Children can discover the value of π as the constant ratio between the circumference and the diameter of the circle.
- The children know the formula of the circumference calculation of the circle and they can apply it.

Please note the following:
- Link this lesson to *the curriculum*
- Build up the lesson logically and take into account the *general professional didactic principles*.
- Indicate *which course material* you will use and why.
- Provide information about the *classroom organization*.
- Explain how you are going to use the *blackboard and/or write a board plan*.
- Specify how to deal with the fast workers i.e., *differentiation*

***

Low complex task (Question 2): The students are working on the application exercises on calculating the circumference of the circle. The teacher gives weaker students tools to solve the exercises. Is this an example of tempo or level *differentiation*? Explain.

Example of procedural information

All pupils should receive the same instruction and the differentiation. Accordingly, you should differentiate on:
- the number of assignments and the pace at which the pupils carry out the tasks (differentiation on tempo)
- the degree of difficulty of the exercise tasks, the use of tools for solving exercises (level differentiation)
- the extra support of the teacher

Read more about differentiation in the classroom (*link to supportive information*).

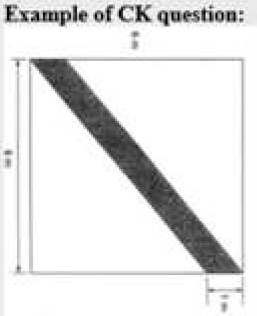
Figure 5.1. Highly complex task, question of the low complex task and an example of the procedural information

Students' prior knowledge

Information about students' prior knowledge was gathered in the first semester during their examination. Students were tested on their knowledge of PCK (*mean* = 63.5%, *SD* = 19.7) and CK (*mean* = 72.2%, *SD* = 27.8). The content was (teaching) mathematics in general and geometry in particular. Examples of test-items can be found in *Figure 2*. All tests were corrected by the instructor of the course unit. We have no insight into the prior knowledge of one student who participated in the study, which means that we can include an indicator of prior knowledge for 14 students in the analysis.

Example of a PCK question:
Which arithmetic skills are learned by means of these exercises?
0,1 x 4180 =
4008 : 8 =

Example of CK question:



Calculate the area of the footpath through our square garden.

Figure 5.2.: Example questions of the prior knowledge test

Subjective measurements

For the measurement of cognitive load a validated instrument developed by Leppink et al. (2013) was used for the measurement of intrinsic, extraneous, and germane load. The questionnaire was translated into the specific context of the present study as shown in Table 1. The questionnaire consisted of a 7-point Likert scale (i.e. ranging from “totally disagree” to “totally agree”). Reliability was determined through Cronbach’s α in order to investigate the overall consistency of the constructs (Schreiber, Nora, Stage, Barlow & King, 2006). Confirmatory factor analysis (CFA) was not conducted due to the small sample size, but former research has validated the questionnaire and has proven that the questionnaire is reliable (Leppink et al., 2013). Additionally, the Paas’s nine-point mental effort rating scale was added to the questionnaire (Paas, 1992).

Table 5.1.

Survey items and reliability of the constructs

	High complex α	Low complex α
ICL_1: The topics covered in this task were very complex	.69	.83
ICL_2: The task covered formulas that I perceived as very complex		
ICL_3: The task covered concepts and definitions that I perceived as very complex		
ECL_1:The instructions during the task were very unclear	.69	.71
ECL_2: The instructions were full of unclear language		
ECL_3:The instructions were, in terms of learning, very ineffective		
GCL_1:The task really enhanced my understanding of the topics covered	.85	.75
GCL_2:The task really enhanced my knowledge and understanding of the topic		

**ICL = intrinsic cognitive load; ECL = extraneous cognitive load: GCL = germane cognitive load*

Physiological data

To measure physiological data including EDA and ST, 15 students were monitored with wrist-worn wearables as shown in Figure 2. These wearables were able to sense GSR with a high dynamic range (.05-20μS) at the lower side of the wrist and the output was accurate within a frame of approximately 1 second. ST was acquired at the upper side of the wrist at a frequency of 32Hz and the output was accurate within a frame of approximately 1 second at 0.1 °C. Before analyzing the physiological data, a number of procedures were carried out. Firstly, a Confidence Indicator (CI), with values ranging from 0 to 1, monitors whether the sensor is correctly attached to the body. Values of CI lower than .80 were ignored as this indicates the low quality of the data due to incorrect sensor attachment (+/- .01% per individual). Secondly, a visual analysis of the signal was conducted for both EDA and ST. Artefacts were removed 20*s* before and after the artefact and an interpolation over the gap was performed. Thirdly, large differences in skin conductance among individuals can occur (Yousoof & Sapiyan, 2013). Therefore, to counteract the variation between subjects, the EDA and ST data of each individual participant were standardized, bringing the mean of each signal to 0 and its variance to 1. Fourthly, time-domain features were analyzed, and mean EDA and ST were calculated as shown in *Figure 3*.

$$\boldsymbol{Mean\ EDA(s,t)} = \frac{\sum_t \boldsymbol{Standardized\_EDA(s,t)}}{\boldsymbol{r}} \qquad \boldsymbol{Mean\ ST(s,t)} = \frac{\sum_t \boldsymbol{Standardized\_ST(s,t)}}{\boldsymbol{r}}$$

**s = subject/ t = task/ r = time-on-task*

Figure 5.3. Standardized mean EDA and ST

Log-data

Log-data was retrieved from the Moodle Learning Management System (LMS). The LMS-system automatically keeps tracks of user activity (i.e. every min) and session. Log-data was divided into several events, namely: (1) start the task; reading instructions, (2) writing an answer, (3) consultation of support and, (4) submission; reviewing the answer.

Analysis

This study first investigated the differences between a high and low complex task for both the subjective measurements and physiological data (i.e. RQ1, RQ2). Therefore, both subjective measurements and physiological data were tested on the normality assumption. Results of the Shapiro-Wilk tests reveal that both subjective measurements and physiological measurements were normally distributed. As we were interested in the mean differences between the high and low complex task of both the self-reported and physiological data, controlled for prior knowledge (i.e., both PCK and CK), order effect (see section 3.1), we conducted a Linear Mixed Model (LMM) incorporating PCK, CK, and order as fixed factors and measurement time as a repeated measure (two-level for RQ1 and three-level for RQ3). When conducting LMM, the Restricted Maximum Likelihood Method (REML) was applied (Baayen, Davidson & Bates, 2008). Based on findings of RQ1 and RQ2, this study investigated the individual differences in the self-reported data of cognitive load for a high and low complex task, and how this relates to individual differences in physiological data (RQ3). Cohen's d was calculated when differences were significant in order to have insight into the effect sizes (LeCroy & Krysik, 2007). A bivariate correlation analysis was conducted in order to find relationships between physiological data and subjective measurements of cognitive load. Fourthly, as the advantage of physiological data is that it is measured continuously, this study investigated whether there are relationships between specific events (i.e. consultation of support) based on log-data and peaks (i.e. spontaneous fluctuations per *s*) of EDA and drops of ST (i.e. RQ4). Given the small sample size, the analysis more descriptive.

5.4. Results

Preliminary analysis

Descriptive statistics of the subjective measurements as shown in Table 2 reveal that students reported on average higher intrinsic load, extraneous load, and mental effort during the highly complex task in comparison with the low complex task. Results furthermore indicate that students reported higher germane load during the low complex task which was expected.

Table 5.2.

Descriptive statistics

	High complex task	*Low complex task*
Cognitive load	*Mean (SD)*	*Mean (SD)*
*Intrinsic load	5.62 *(.97)*	4.78 *(.94)*
*Extraneous load	5.13 *(.84)*	5.31*(1.13)*
*Germane load	3.33 *(2.26)*	3.60 *(1.88)*
**Mental effort	6.47 *(.92)*	4.93 *(1.10)*

7-point Likert scale/9-point Likert scale*

In order to investigate differences in the perceived cognitive load and mental effort (i.e. RQ1), LMM was conducted incorporating PCK, CK, order effect as fixed factors and time as a two-level repeated measurement. Pairwise comparison of the different measurements of intrinsic load, extraneous load, germane load and mental effort are indicated in Table 3. Results reveal that intrinsic load differed significantly across phases. $F(1,13) = 6.43$, $p = .03$. Pairwise comparison reveals that intrinsic load was significantly higher ($M = .86$, $p = .03$) during the high complex task with Cohen's $d = .88$. When investigating the fixed factors, there was no significant effect of both PCK, $F(1,10) = .05$, $p = .82$ and CK, $F(1,10) = .43$, $p = .53$. Moreover, no significant order effect was found $F(1,10) = 12$, $p = 74$. As expected, results reveal no significant difference for extraneous load across phases $F(1,13) = 17$, $p = .69$. Pairwise comparison reveals no significant mean difference ($M = -.05$, $p = .90$) between the high and low complex task for extraneous load. Results of the fixed effects reveal no significant effect of PCK $F(1,10) = .04$, $p = .84$, CK $F(1,10) = .17$, $p = .69$, and order $F(1,10) = 1.58$, $p = .24$. Results for germane load indicate no significant differences across phases $F(1,13) = 1.21$, $p = .29$. Pairwise comparison reveals no significant mean difference for germane load ($M = -.18$, $p = .29$) between the high and low complex task. Results of the fixed effects indicate no significant effects for PCK, $F(1,10) = .00$, $p = .96$ and CK, $F(1,11) = .01$, $p = .93$. Moreover, no order effect was found, $F(1,10) = 1.39$, $p = .2$. Finally, results revealed that mental effort was different across phases. Mean difference of mental effort between the high and low complex task was significant ($M = 1.43$, $p = 00$) in the predicted direction with Cohen's $d = 1.52$. No significant effects of PCK, $F(1,11) = 2.39$, $p = .15$ and CK, $F(1,11) = 2.84$, $p = .12$. Additionally, no order effect, $F(1,10) = .27$, $p = 62$ was found.

Table 5.3.

Pairwise comparison of subjective measurements controlled for prior knowledge (i.e. PCK, CK) and order effect

high-low complex	Mean difference	*BCa*	*p*
Intrinsic load	.86	[.13, 1.59]	.03*
Extraneous load	-.05	[-.79, .89]	.90
Germane load	-.18	[-.53, .17]	.29
Mental effort	1.43	[.65, 2.20]	.00*

**significant at the .05 level **significant at the .01 level; BCa = 95% Confidence interval for Difference*

Descriptive statistics of the physiological data can be found in Table 4. Mean EDA is lower during the high complex task compared to the low complex task. Mean ST is lower during the high complex task.

Table 5.4.

Descriptive statistics of the standardized physiological data

Physiological data	Baseline measurement Mean (*SD*)	High complex task Mean (*SD*)	Low complex task Mean (*SD*)
Mean EDA	-.58 (*.60*)	.09 (*.45*)	.45 (*.86*)
Mean ST	1.25 (*.86*)	.35 (*.38*)	.49 (*.87*)

In order to investigate the differences of physiological data between the baseline measurement, high and low complex task (i.e. RQ2), LMM was conducted incorporating PCK, CK, order effect as fixed factors and time as a three-level repeated measurement. Results indicate that differences were found for mean EDA across the different phases $F(2,26) = 6.56$, $p = .01$. Pairwise comparison of the different measurements of mean EDA is indicated in Table 5. Results of pairwise comparison reveal that the mean difference between the baseline measurement and high complex task phase is significant in the predicted direction ($M = -.60$, $p = .05$) with Cohen's $d = .19$. Moreover, the mean difference is significant between the baseline measurement and the low complex task ($M = -1.05$, $p = .00$) with Cohen's $d = .14$. Results reveal that no significant mean difference was found between the high and low complex task ($M = -.45$, $p = .14$). Moreover, the mean difference was in the unexpected direction. When investigating the fixed factors, there was a non-significant main effect of both PCK $F(1,10) = .18$, $p = .68$ and CK $F(1,10) = .81$, $p = .36$. Additionally, there was a significant effect of order $F(1,10) = 7.62$, $p = .02$, which indicates an order effect.

No significant differences were found for mean ST across the different measurements, $F(2,26) =.16$, $p = .85$. Pairwise comparison reveals no significant mean differences between baseline measurement and the high complex task ($M = 1.02$, $p = .61$), baseline measurement and the low complex task (M = .87, p = .66), and between the high and low complex task ($M = -.15$, $p = .94$). Nonetheless, all mean differences were in the expected direction. When investigating the fixed effects, there was a non-significant main effect of both PCK $F(1,10) = .00$, $p = .97$ and CK $F(1,10) = .12$, $p = .74$. Additionally, there was no significant order effect, $F(1,10) = .45$, $p = 52$.

Table 5.5.

Pairwise comparison of physiological data controlled for prior knowledge and order

Physiological data: phase	Mean difference	*BCa*	*p*
Mean electrodermal activity			
Pair 1: Baseline – high complex	-.60	[-1.20, .00]	.05*
Pair 2: Baseline- low complex	-1.05	[-1.65, -.45]	.00**
Pair 3: High complex- low complex	-.45	[-1.05, .15]	.14
Mean skin temperature			
Pair 1: Baseline – high complex	1.02	[-2.98, 5.01]	.61
Pair 2: Baseline – low complex	.87	[-3.13, 4.87]	.66
Pair 3: High complex- low complex	-.15	[-4.14, 3.85]	.94

**significant at the .05 level **significant at the .01 level; BCa = 95% Confidence interval for Difference*

Results of RQ1 reveal significant differences in perceived intrinsic load and mental effort. RQ3 investigates the relationship between the individual differences of intrinsic load, mental effort, and physiological data. Results are displayed in Table 6 and reveal that mental effort is significantly positively correlated with mean EDA (r = .58, p = .03) for the highly complex task. Nevertheless, no significant positive correlation was found between mean EDA and mental effort for the low complex task. No significant results were found for ST.

Table 5.6.

Correlations between standardized physiological data and subjective measurements for the highly complex and low complex task.

	High complex task				Low complex task			
	Mean EDA		Mean ST		Mean EDA		Mean ST	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Intrinsic load	.12	.34	-.04	.44	.16	.29	-.03	.46
Mental effort	.58	.03*	.33	.12	.12	.34	-.01	.48

*** correlation is significant at the .01 level; * correlation is significant at the .05 level*

In the final RQ4, this study investigates the relationship between physiological data and specific events retrieved from log-data and EDA peaks. An example of such relationships is shown in *Figure 4*. Table 7 gives an overview of the number of relationships between specific events and EDA peaks. In contrast to EDA, no conclusive relationships were found between ST (i.e. drops) and specific events. ST for most participants increased throughout the intervention as illustrated in *Figure 5*.

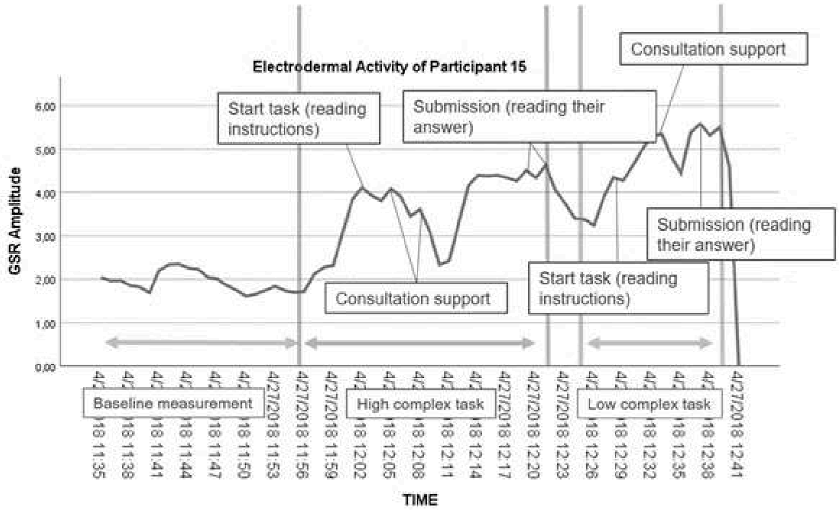


Figure 5.4. Electrodermal activity related to log-data of participant 15

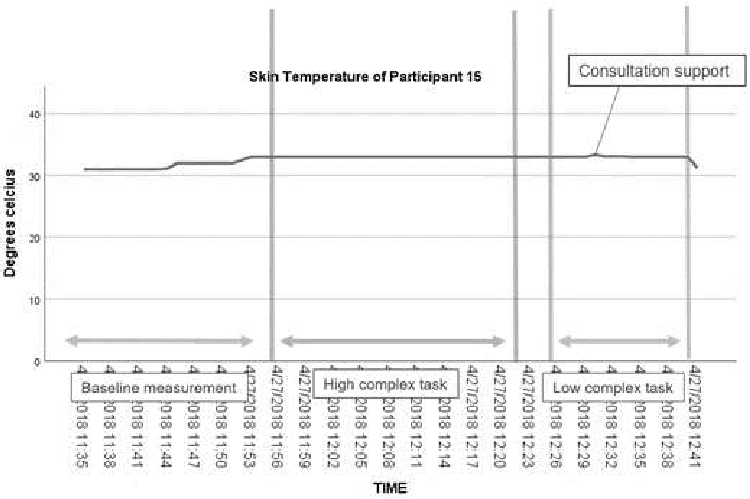


Figure 5.5: Skin temperature related to log-data of participant 15

Table 5.7.

The relationship between specific events and EDA peaks

Events	High complex	Low complex
Start the task (reading instructions)	7	2
Writing an answer	8	2
Consultation support	8	3
Submission (reviewing the answer)	6	14

5.5. Discussion

This study attempted to *firstly* investigate the difference of subjective measurements of cognitive load between a high and low complex task (i.e. *RQ1*). Results reveal that the students indicate higher perceived *intrinsic load* for the high complex task when compared with the low complex task. This indicates that the manipulation of complexity based on element interactivity was successful. Additionally, students indicated that the perceived *mental effort* was higher during the high complex task. Effect sizes of both intrinsic load and mental effort were high (>.80) indicating that the manipulation of complexity had an impact (LeCroy & Krysik, 2007). This reveals that students invested more mental effort into solving the high complex task in order to maintain performance at a constant level (Paas et al., 2005). This is also in line with CLT, since the high complex task was high in element interactivity and possibly required a lot of cognitive resources (Van Merriënboer & Sweller, 2005). No significant differences were found for *extraneous load* between both tasks. This finding was expected as the instructions for both tasks were of the same level of difficulty. Additionally, no significant differences were found for *germane load,* indicating that both tasks enhanced students' understanding of the content at a similar level. This was in line with our expectations as the content and available support of both tasks was the same.

Secondly, this study aimed at investigating whether we can use physiological data to distinguish between the two complexity levels of the task. When investigating *mean EDA*, results reveal that significant differences were found between both tasks and the baseline measurement. These findings indicate that both tasks result in a higher mean EDA. Nevertheless, effect sizes were very small (< .20), indicating that task complexity only had a minimal impact on mean EDA (LeCroy & Krysik, 2007). Moreover, no significant differences were found for mean EDA between the high and low complex. These results are in line with the findings of the study of Haapalainen et al. (2010), which also revealed no significant differences for EDA between six tasks of different levels of difficulty. Moreover, against expectations, descriptive statistics reveal that mean EDA was higher during the low complex task, when compared with the high complex task. These unexpected findings may be induced by the order effect. This order effect may reduce a clear difference between the EDA during the high and low complex task. Moreover, visual analysis reveals that for the majority of all participants, skin conductance rises throughout the intervention (i.e., drift). Since, more participants had the low complex at the end, this might indicate that results are biased by drift. This indicates the need for the current study to also examine EDA peaks as these peaks are not affected by drift (*RQ4*).

When investigating *mean ST* no significant mean differences were found for mean ST across all different phases. Nevertheless, descriptive statistics reveal that ST was higher during the baseline measurement period. Moreover, ST was higher during the low complex task compared with the high complex task. This could indicate that ST is related to task complexity as research indicated that ST declines relative to a trigger event (Ikehara & Crosby, 2005). Current findings indicate that mean EDA and mean ST might be indicators of changes of cognitive load, but cannot be used to detect differences in task complexity. Nevertheless, there is no clear link between ST and cognitive load. Accordingly, correlations between individual differences in the perceived intrinsic load, mental effort and, physiological data for a high and low complex task are investigated (*RQ3*).

A *third aim* of this study was to investigate whether we can relate subjective measures of the perceived intrinsic load and mental effort (i.e. based on findings of *RQ1*) with physiological data (i.e. mean EDA and ST) during a high and low complex task. Findings reveal that mental effort positively correlates with *mean EDA* for the high complex intervention. Nevertheless, we did not find a significant correlation between mean EDA and the low complex intervention Results might also be influenced by the fact that skin conductance was rising throughout the intervention, accompanied with the fact that more students completed the low complex task at the end. No significant correlations between *mean ST* and self-reported data were found. This finding could be due to the fact that ST shows a very slow rise and decline in temperature change relative to the trigger event. Therefore, it might be difficult to relate ST to self-reports (Ikehara & Crosby, 2005). Since there seems to be a relationship between EDA and mental effort and since ST drops can be related to specific events, we investigated the relationship between physiological data and learning behavior retrieved from log-data.

In order to investigate the relationship between physiological data and learning behavior. Log-data was investigated and divided into four main events, namely, reading instructions, writing an answer, consulting support and reviewing the answer. Results reveal that there seems to be a relationship between specific learning behaviour and EDA peaks. Moreover, results reveal that more peaks were registered during the high complex task, when compared with the low complex task, which indicates a different result compared to *RQ2*. When investigating the intensity of the peaks, findings reveal that the peaks that are related to the events 'submission' are more intense. This might explain, besides the occurrence of drift, why mean EDA was higher during the low complex task. Possibly, results may have been influenced by the fact that the low complex task was presented as a test-format, which might have induced more intensive peaks when students submitted their task. When investigating relations between peaks and events it seems that during the high complex task, peaks are more related to cognitive processes (e.g., reading instructions, consulting support and writing) when compared with the low complex task (e.g., submission). For instance, when investigating the event 'consultation of support more in detail, peaks were related to students (*N* = 4) watching a video that explains the circumference of a circle.
This is line with previous research indicating that GSR responses are associated with effortful cognitive processing during multimedia learning (Antonietti, Colombo & Di Nuzzo, 2015). Additionally, hardly any peaks were found for the low complex task during writing, which is in line with the study of Mudrick, Taub, Azevedo, Price & Lester (2017). Mudrick et al. (2017) investigated multimedia learning and indicated that the lowest amount of GSR responses were retrieved when answering multiple-choice questions, suggesting that this might require less cognitive processing. This finding is also in line with the study of Hoogerheide et al. (2018) indicating that mean EDA was significantly lower during the solving process of a practice problem, when compared with teaching a practice problem in an authentic learning situation. These exploratory findings indicate that the intensity of EDA signals might be more related to the type of learning activities. In line with previous findings of *RQ2* and *RQ3*, no conclusive results were found for ST. Nevertheless, on the basis of data visualization of all students we could see that for the largest number of participants (i.e., 8 students), ST is lower during the high complex task, which is in line with findings of *RQ2.*

Limitations and further research

Despite the merits of the study in terms of indicating that individual differences in the experienced mental effort can indicate individual differences in EDA, there are some important limitations that should be mentioned. Firstly, results must be approached carefully as multiple analyses on the same dependent variable were conducted which can increase the chance of committing a Type 1 error (Roth, 1999). Secondly, as we were investigating physiological data, we were obliged to implement a within-subject design. This is a requirement when investigating skin conductance, as skin conductance can vary markedly between individuals (Braithwaite et al., 2013). Nevertheless, the within-subject design had some important disadvantages. Since the same learning materials were taught within both the high and low complex task, students might have learned from the previous task and therefore perceived the high complex task as less difficult. This is turn might have influenced skin conductance and skin temperature, and might be the reason why there was no clear difference between the high and low complex task. This problem can be addressed in future studies by addressing different topics. Moreover, future studies should offer more different tasks of different levels of complexity, and also create more conditions in order to increase the number of measurements. This could provide a better understanding of possible correlations between mental effort and mean EDA.

A third important limitation, when investigating skin conductance is drift, a continuous increase of the intensity of the signal. It is important to distinguish drift from important shifts in real tonic processes (Braithwaite et al., 2013). Nevertheless, this distinction between drift and real tonic processes is not always entirely clear. This emphasizes the need of an accurate baseline measurement. The baseline measurement in the current study could be optimized by letting students rest. Given the small sample size we decided not to remove data of participants. Instead, in this study we have additionally investigated the peaks of skin conductance (as these are no subject of drift) and related them to specific events in the learning environment. Nevertheless, it can be advisable to remove the data of participants on the basis of drift in larger datasets. Moreover, a larger sample size would also allow us to investigate patterns between EDA peaks and specific events in the learning environment (e.g. reading instructions) while using quantitative methods. Finally, as the study did not take place in a lab setting but in the classroom of the students, a lot of confounding factors unrelated to cognitive load may cause clouds in the data such as a lecturer entering the classroom and students leaving the classroom when finished.

These events are likely to degrade the accuracy of cognitive load measurement by GSR (i.e. EDA). Nevertheless, the ecological valid setting also has advantages such as authenticity of the results (Schmuckler, 2001). Moreover, as the subject matter was part of students' training program, students were encouraged to thoroughly solve the tasks, which is reflected in the task performance

5.6. Conclusion

This study attempted to firstly investigate the difference of subjective measures of cognitive load and physiological data (i.e. mean EDA and ST) between a high and low complex task in an ecologically valid setting. Students indicated that they perceived higher intrinsic load during the high complex task and that the high complex task required more mental effort. This indicates that task complexity can be manipulated based on element interactivity. Nevertheless, complexity was not reflected by differences in physiological data (i.e. mean EDA and ST). Accordingly, in the next phase this study investigated correlations between perceived intrinsic load, mental effort and physiological data. Results revealed a positive correlation between mean EDA and mental effort during the high complex task. Nevertheless, no significant correlations were found for the low complex task. Preliminary results of more descriptive analysis showed that peaks of EDA during the high complex task were more frequently related to cognitive processes when compared with the low complex task (i.e. submitting the task). The latter finding might explain the significant relationship between mental effort and mean EDA. Future research should replicate similar studies while using larger sample sizes to verify these findings. Additionally, the relationship between EDA and the type of learning behavior (i.e. retrieved from log-data) should not be overlooked.

Chapter Six

Physiological data: a promising avenue to detect cognitive (over)load?

This chapter was published in a special section on MMLA as: Larmuseau, C., Cornelis, J., Lancieri, L., Desmet, P., & Depaepe, F. (2020). Multimodal learning analytics to investigate cognitive load during online problem solving. *British Journal of Educational Technology, 51,* 1548-1562. https://doi.org/10.1111/bjet.12958

6. Physiological data: a promising avenue to detect cognitive (over)load?

6.1. Introduction

In the search for a better understanding and the support of learning, new manners of data collection such as sensing technology are explored to capture multimodal data unobtrusively in ecologically valid learning environments (Spikol & Cukurova, 2019). In the current study we used physiological data to measure cognitive load during the online problem-solving process. Cognitive load was defined based on the Cognitive Load Theory (CLT) introduced by Sweller (1994). CLT indicates that cognitive load can be induced by both *intrinsic* and *extraneous* load (Sweller, 1994). The level of *intrinsic load* is determined by the amount of element interactivity and their interrelationships that need to be mastered by the learner. *Extraneous load* is mainly imposed by instructional procedures that induce unnecessary working memory load (Sweller, 2010). By monitoring task complexity and instructional support, personalized online courses can be developed to optimize cognitive load. In view of optimizing cognitive load, it is important to accurately measure cognitive load during the online problem-solving process. Former studies used physiological measurements such as GSR, ST, HR and HRV to investigate cognitive load (Cranford, Tiettmeyer, Chuprinko, Jordan & Grove, 2014; Haapalainen et al., 2010; Larmuseau et al., 2019; Nourbakhs et al., 2012). Despite the merits of these studies our current understanding of the association between physiological data and cognitive load is characterized by at least three limitations. Firstly, most studies did not systematically manipulate cognitive load based on insights from CLT (Morton et al., 2019). Secondly, the majority of studies did not combine physiological data with self-reports and task performance (Dindar, Malmberg, Järvelä, Haataja, & Kirschner, 2019). Thirdly, there is no unambiguous answer to which physiological features are best in assessing cognitive load (Larmuseau et al., 2019).

To meet these limitations, the current study firstly experimentally manipulated the intrinsic load and the extraneous load by respectively varying the difficulty level of statistical exercises and the instructional support. More particularly, four sets of exercises were developed that differed on these two dimensions. In addition, students also received a computer-based operation span test (OSPAN) and participated in baseline measurements. OSPAN was used as verification of high cognitive load, whereas the baseline measurement was a verification of low cognitive load (Yuan et al., 2006). We examined differences between the four sets, OSPAN and baseline measurement, in view of the physiological data. Secondly, we combined physiological data with self-reported cognitive load and task performance to investigate how much variance in these variables was explained by physiological data. Thirdly, we investigated which physiological features were important for distinguishing high and low cognitive load.

6.2. Theoretical background

Cognitive Load Theory

CLT was introduced by Sweller (1994) to consider the instructional implications of characteristics of human cognitive architecture with a special focus on the limited capacity of the working memory. Cognitive load is defined as working memory load, which is determined by the working memory resources required by a learner, for performing a cognitive task (Kalyuga & Singh, 2016). A major goal of CLT is to optimally manage cognitive load, since both overload and underload can lead to substandard performance (Chen et al., 2016). CLT distinguishes between intrinsic and extraneous load (Paas et al., 2010). *Intrinsic load* is determined by the level of element interactivity. In the case of higher element interactivity, the learning material is more complex requiring more intrinsic processing from learners' working memory for coordinating and integrating the learning material. *Extraneous load* refers to extraneous processing that does not contribute or even obstructs learning due to unnecessary mental demands (Sweller et al., 2019). Instructional design techniques can reduce extraneous load by preventing learners to pay attention to irrelevant elements. For instance, instructional support can contribute to reducing this extraneous load by offering hints to tackle the learning tasks (Cierniak et al., 2009; Sweller, 2010). In view of optimal learning and performance, an optimal level of cognitive load is desirable (Sweller et al., 2019). Therefore, it is important to accurately assess cognitive load during the online problem-solving process to detect high complex learning material or suboptimal instructional procedures. Accordingly, personalized online learning environments can be developed that are adjusted to the learners' working capacity levels.

Measurement of cognitive load

A typical approach for measuring cognitive load is through unidimensional rating scales (Chen et al., 2016). In this respect, a frequently used rating scale is the Paas' (1992) nine-point mental effort rating scale (Chen et al., 2016). This scale requires learners to rate their mental effort immediately after completing a task (Paas, 1992). This scale can be used as an index of overall cognitive load (Chen et al., 2016). Despite the merits of this rating scale, such as a quick administration of the perceived cognitive load, some limitations should be mentioned. Firstly, this scale requires learners to introspect their cognitive processes which can induce biased results (Boekaerts, 2017). Secondly, these scales are obtrusive as they interrupt task flow. Thirdly, those rating scales do not easily capture variations in load over time (Chen et al., 2016). In contrast to self-reported data, physiological data can be measured at a high frequency and with a high precision during online problem solving (Di Mitri; Schneider, Specht & Drachsler, 2018). Given the close relationship between cognitive load and neural systems, the physiological approach is seen as a promising avenue to assess cognitive load (Chen et al., 2016). In the next section, we will explain how physiological data such as GSR, HR(V) and ST, can be unobtrusively measured by means of wrist-worn wearables and patches.

Galvanic skin response (GSR) or Electrodermal Activity (EDA)

GSR, also known as skin conductance or electrodermal activity (EDA) refers to the variation of the electrical properties of the skin in response to sweat secretion (Benedek & Kaernbach, 2010). The time series of skin conductance can be characterized by a slowly varying tonic activity (i.e. skin conductance level) and a fast varying phasic activity (i.e. skin conductance responses; Braithwaite et al., 2015). The bulk of skin conductance literature mainly reports associations with stress (Braithwaite et al., 2015; Smets et al., 2018a). Nonetheless, more and more researchers also investigated the relationship between GSR and cognitive load (Chen et al., 2016). Nourbakhs et al. (2012) captured GSR data from learners conducting arithmetic and reading tasks. The tasks differed in difficulty level, which relates to intrinsic load. Resp. four and three difficulty levels were distinguished for the arithmetic and reading tasks. Results of ANOVA of 13 and 16 participants (arithmetic and reading tasks) indicated that GSR significantly differed between task difficulty levels. By contrast, Shi et al. (2007) investigated 11 subjects when dealing with four tasks that differ in level of difficulty, but the results revealed insignificant differences across task difficulty levels for GSR. Similarly Larmuseau et al. (2019) investigated GSR for tasks that differed in terms of element interactivity. Results of 15 participants indicated no noticeable difference in GSR data between a high and low element interactivity task. Nonetheless, significant differences were found between GSR during the baseline measurement (i.e. low cognitive processing) and during high complex problem-solving. Overall, studies indicate that GSR increases when cognitive load increases.

Heart Rate (HR) and Heart Rate Variability (HRV)

HR and HRV can be measured by a noninvasive electrocardiographic (ECG) method. HR averages the number of beats per minute, whereas HRV indicates small changes in the intervals between successive heartbeats. The majority of the studies have revealed that HR and HRV can be used to measure stress (Kim, Cheon, Bai, Lee & Koo, 2018; Smets et al., 2018a). Other studies have also associated HR and HRV with cognitive demands, in which they have shown that an increase in HR and a decrease in HRV indicates higher cognitive load. For instance, Taelman, Vandeput, Vleminckx, and Spaepen (2011) collected ECG data of 43 undergraduates in a laboratory experiment during a high cognitive load task (doing complex arithmetic exercises) and a low CL task (watching a relaxing movie). Results of pairwise comparison, reveal that HRV was significantly higher during the rest phase when compared with the high cognitive load task. Additionally, Cranford et al. (2014) measured CL through HR in the context of chemistry. Findings of 12 participants suggested that problems that were intentionally designed to induce higher cognitive load resulted in a larger increase of HR, compared to problems designed to induce lower cognitive load. Finally, in the study of Brouwer, Hogervorst, Holewijn, and van Erp (2014), 35 participants solved different difficulty levels of the n-back task (i.e. working memory capacity test) while recording HR and HRV. No significant effects were observed for HR and HRV, but trends in the data revealed that HR varied as a consequence of task difficulty. In summary, we can assume that HR and HRV are promising in detecting changes in cognitive load.

Skin Temperature

Previous studies also measured ST to indicate stress. Stress can induce peripheral vasoconstriction which causes a rapid, short-term drop in ST. Moreover, stress can also cause a more delayed skin warming, providing two opportunities to quantify stress (Herborn et al., 2015; Karthikeyan et al., 2012; Smets et al., 2018). Little research has used ST to assess cognitive load. As an exception, Haapalainen et al. (2010) collected data from multiple sensors, one tracking ST, in view of detecting cognitive load. A total number of 20 subjects had to solve six tasks that differed in complexity. Using personalized machine learning techniques (i.e., Naïve Bayes Classifier) to assess cognitive load, they concluded that ST can be used to distinguish between low and high complex tasks. By contrast, the study of Larmuseau et al. (2019) observed no significant differences in ST between low and high complex tasks. In general, studies that used ST to investigate cognitive load remain scarce. Consequently, it is unclear whether ST can be used to detect cognitive load.

Shortcomings and research questions

Overall, there are some limitations in the current research field. Firstly, not all studies systematically manipulated cognitive load according to CLT (Morton et al., 2019). Secondly, the majority of studies did not combine physiological data with self-reported cognitive load (Dindar et al., 2019; Larmuseau et al., 2019). In addition to linking physiological data with self-reported cognitive load it should be interesting to link physiological data with task performance across the different sets of exercises, as previous studies indicated that physiological adjustments ensure that task performance remains the same across different levels of complexity (Brouwer et al., 2014; Iani, Gopher & Lavie, 2002). Thirdly, it remains unclear which features are most indicative for assessing cognitive load. Against this background, the following research questions are formulated:

- ***RQ1:*** Do the different conditions (i.e. sets of exercises, OSPAN and baseline measurement) result in differences in physiological data?
- ***RQ2a****:* Can self-reported cognitive load be explained in terms of physiological data across the different sets of exercises?
- ***RQ2b****:* Can task performance be explained in terms of physiological data across the different sets of exercises?
- ***RQ3:*** Which physiological features are important in assessing cognitive load?

6.3. Method

Participants and Study design

Participants were 67 adolescents (67.2% female and 32.8% male). The average age was 19.37. Of these 67 participants, 21 were in there last year secondary education and 46 were in the first two years of higher education. Participation was voluntary and all participants signed an informed consent before participation. A requirement for participation was that participants had been recently introduced to the theory on probabilistic reasoning in statistics. A within-subject design was used in which participants solved four different sets of exercises in a randomized order. The intervention is illustrated in *Figure 1*. Participants started with OSPAN. Afterwards, baseline measurements were conducted where students watched a relaxing movie with headphones. Additionally, participants randomly received (to counteract a sequence effect) four sets of exercises (section 3.4.) within the Moodle LMS. All answers were completed online and students were allowed to use a calculator on their computer to conduct calculations if they felt that this was useful. No information about the correctness of the answers was provided. After each set of exercises, participants had to indicate their perceived cognitive load (Paas, 1992) (Q*).

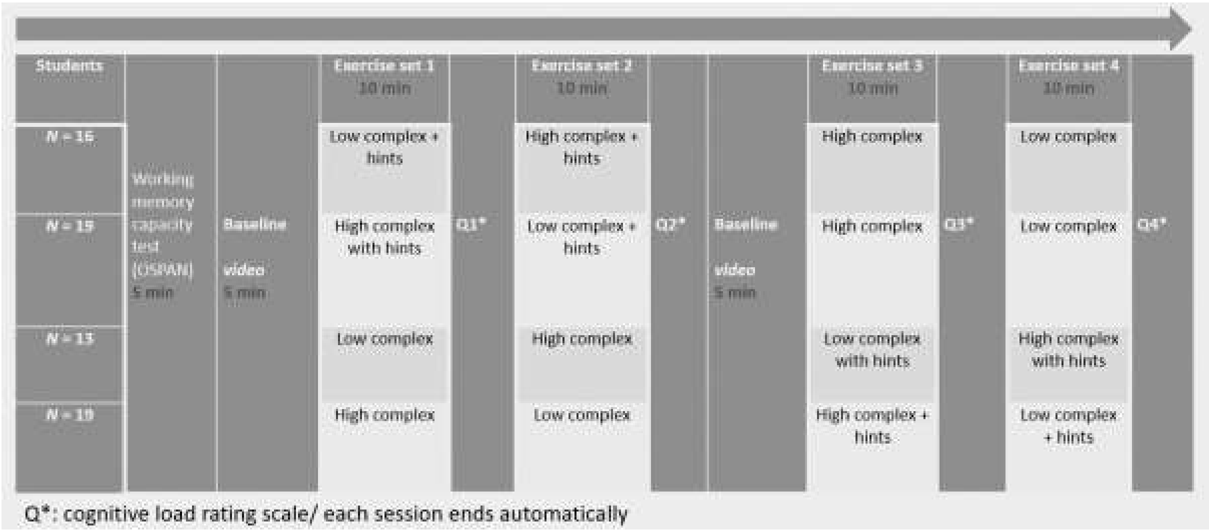


Figure 6.1.: The study design

Physiological recordings

During the sets of exercises physiological data were measured by two wearables developed by Imec as indicated in *Figure 2*. The first wearable was a chest patch recording ECG at a sampling rate of 256 *Hz* and providing information about HR(V). The second wearable was a wrist-worn device (worn on the non-writing hand) measuring GSR with a high dynamic range (.05-20 μS) at the lower side of the wrist. The output was accurate within a frame of approximately 1 second. ST was acquired at the upper side of the wrist at a frequency of 1 *Hz*. The output was accurate within a frame of approximately 1 second at 0.1°C. Both wearables measured the magnitude of acceleration. Using high-quality physiological signals, 19 features were calculated (8 GSR features, 4 ST features and 7 ECG features) in a window of 1 min. There was drop-out due to the malfunctioning of some wearables (e.g. wearables that went into standby, or patches that could not measure due to chest hair, etc.), resulting in data from 51 wrist-worn wearables (GSR and ST) and 48 ECG patches (HR and HRV).

Figure 6.2.: The Imec chest patch and Chillband

The computer-based operation span test (OSPAN)

A computer-based operation span test (OSPAN; Stone & Towse, 2015) was used in which each participant was shown a number for 1s on a screen that he/she had to remember in the correct order. After each number, participants were shown a mathematical operation and had to decide on the correctness of the given answer. Accordingly, for every storage element (number to remember) there was a processing phase (a mathematical operation) immediately succeeding it. In total there were three experimental trails of each span size (number of digits to be remembered: 2-6). Since this dual-task requires both storage and the processing of the information, it is effortful and demanding of the working memory (Yuan et al., 2006). Consequently, this test can be used as a control intervention for high cognitive load.

Exercises and task performance

The content of the exercises was probabilistic reasoning in statistics. The four sets of exercises that were developed differed from each other on two dimensions. The first dimension was related to intrinsic load.

The complexity of the exercises was manipulated on the basis of element interactivity (Sweller, 2010). As illustrated in Figure 3, solving the low complex set of exercises consisted of applying one procedure or rule while the high complex exercise series consisted of several solution steps consisting of different elements (elements are in bold). The second dimension was related to extraneous load and was manipulated by providing step-by-step instructions (i.e. hints) that were offered just-in-time and could be consulted voluntarily (in italics in *Figure 3*). Consequently, we had four sets of exercises: (1) high complex with hints, (2) high complex without hints, (3) low complex with hints and, (4) low complex without hints. By systematically manipulating cognitive load we aimed at detecting differences in cognitive load through physiological data. Task performance was retrieved from the total number of exercises that was correctly solved per session. All exercises had a precisely correct answer and were scored as correct (1) or incorrect (0).

Low element interactivity	**High element interactivity**
Exercise: In a class, 5 children have to read. In how many different ways can you create a schedule with the order in which they have to read aloud? *Hints provided* ***(one element):*** *- Apply* ***the permutation rule*** *(order is important)*	**Exercise:** You have 8 men and 10 women. You want to form a jury of 12 members. But you do want to have more men than women in the jury. How many unique combinations are there? *Hints provided* ***(three elements):*** - *Apply* ***the combination rule*** *(no strict order and no repetition)* - *Calculate situation 1 (7 men/5 women) and apply* ***the product rule*** - *Calculate situation 2 (8 men/4 women) and apply the product rule* - ***Add up*** *the total possibilities*

Figure 6.3.: example of an exercise of low and high element interactivity with hints

Analysis

Since features of physiological data might be strongly related, principal component analysis (PCA) was used to reconstruct the original dataset of 19 features. By conducting PCA, 7 features were extracted based on their standardized loadings after varimax rotation, namely: GSR_SCmag, ECG_sdnn, ST_mean, ECG_LF, LF_VLF_Ratio, ST_slope and ECG_meanHR. Detailed information about these features (meaning and calculation) can be found in Table 1, in which the selected features are in bold. The purpose of the standardization was to facilitate individual-difference comparisons and thus to factor out issues (e.g. the thickness of the skin; Braithwaite et al., 2015).

Table 6.1.

Overview of the features of the physiological data (selected features in bold)

Features	*Description*
GALVANIC SKIN RESPONSE	
GSR_SCL	Skin conductance level- average GSR
GSR_SCPh	Phasic skin conductance- signal power of the phasic GSR signal
GSR_SCRR	Skin conductance response rate – number of GSR responses in window divided by the total length of the window (i.e. responses per second)
GSR_SCdiff2	Skin conductance second difference - signal power in second difference from a GSR signal
GSR_SCR	Skin conductance response - number of GSR responses from a GSR signal
GSR_SCmag	Skin conductance magnitude - the sum of the magnitudes of GSR responses
GSR_SCdur	Skin conductance duration - the sum of the duration of GSR responses in seconds
GSR_SCarea	Skin conductance area - the sum of the area of GSR responses in seconds. The area is defined using the triangular method (1/2*s_mag*s_dur)
SKIN TEMPERATURE	
ST_mean	Mean of the skin temperature
ST_median	Median of the skin temperature
ST_std	Standard deviation of the skin temperature
ST_slope	Slope of the skin temperature – slope of a straight line fitted through the data
HEART RATE AND HEART RATE VARIABILITY	
ECG_meanHR	Mean heart rate in window
ECG_sdnn	Standard deviation of the interbeat (RR) intervals
ECG_rmssd	Root mean square of the successive interbeat (RR) differences
ECG_LF	Low frequency signal (power in the 0.04-0.015 Hz band)
ECG_HF	High frequency signal (power in the 0.15-0.4 Hz band)
ECG_HC	Heart coherence (a pattern with nice even waves)
ZLF_VLF_Ratio	The low frequency signal (LF)/ (very low frequency signal (VLF) and high frequency signal (HF))

In view of *RQ1* we investigated the mean differences in physiological data between the four sets of exercises, OSPAN and baseline measurement. Since the first baseline measurement was strongly influenced by the measurement of physiological data during the OSPAN (indicated by visual analysis), we opted to only use the second baseline measurement for analysis. We controlled for the magnitude of acceleration (ACC), as movement can influence physiological signals (Boucsein, 2012). A within-subjects effects of the different interventions by 6*1 repeated measure ANOVA was conducted. As multiple comparisons were conducted, Bonferroni correction was applied (Bretz, Hothorn & Westfall, 2016).

Regarding *RQ2* we investigated whether self-reported cognitive load and task performance can be explained in terms of physiological data across the different sets of exercises. Multilevel Modeling (MLM) was applied to investigate how much variance of the self-reported data and task performance was explained by the physiological data. A first null model included the intrinsic manipulation and the extraneous manipulation. The subsequent analyses had perceived cognitive load and task performance as dependent variables and the selected physiological features (i.e. based on PCA) as predictors. The average variance explained by the predictors (R^2) is presented for each model. We also checked if the model was statistically better after inclusion of physiological data. To maintain statistical power while comparing the models we decided to complete missing data via multiple imputation which resulted in data of 64 students. Creating multiple imputations, as opposed to single imputations accounts for the statistical uncertainty in the imputations and yields accurate standard errors (Azur, Stuart, Frangakis & Leaf, 2012). Accordingly, this method can be used for a large amount of missing values such as +/-30% missing observations in the current study (Fichman & Cummings, 2003).

With respect to *RQ3* we investigated whether cognitive load can be detected through physiological data. Informed by the results of RQ1, we only maintained the two interventions that significantly differed and conducted a binary classification, i.e. baseline measurement (low cognitive load) and OSPAN (high cognitive load). The selected features (based on PCA to avoid overfitting) were incorporated in the machine learning model. Using 128 observations (from 64 students) we trained a logistic regression model in a 10-fold cross-validation approach (Ramasubramanian & Singh, 2016).

6.4. Results

RQ1 investigated the differences between the four sets of exercises, OSPAN, and the baseline measurement in terms of physiological data as indicated in Table 2. First, regarding *GSR* (*N* = 51), no significant differences between the six interventions were observed. Secondly, mean *HR* was significantly lower during the baseline measurement compared to the high complex set, high complex set with hints, low complex set with hints and OSPAN. Furthermore, *mean HR* (*N* = 48) was significantly lower during the four sets of exercises when compared with OSPAN. In terms of *HRV* (*N* = 48), results of SDNN (i.e. standard deviation of the interbeat intervals) reveal that HRV was significantly higher during OSPAN when compared with the baseline measurement. Thirdly, results of *ST* (*N* = 51) reveal that mean ST during OSPAN was significantly lower compared to all other interventions. Similarly, the slope of ST was higher during OSPAN when compared with the remaining interventions.

Table 6.2. Overview of the within-subjects effects of the different interventions

	GSR		HR		HRV						ST			
	GSR_ SCmag		ECG_ meanHR		ECG_ SDNN		ECG_LF		LF_VLF_ Ratio		Mean ST		ST slope	
I-J	*MD*	*SE*	*MD*	*SE*	*MD*	*SE*	*MD*	*SE*	*MD*	*SE*	*MD*	*SE*	*MD*	*SE*
Base - HC	-.03	.17	-.25**	.07	.04	.15	.02	.06	-.30	.17	-.02	.06	-.37	.16
Base - HC + hints	-.14	.17	-.36**	.07	.22	.15	-.02	.06	-.16	.17	.04	.06	-.43	.16
Base - LC	-.06	.17	-.18	.07	.23	.15	.06	.06	-.00	17	-.03	.06	-.10	.16
Base - LC + hints	.12	.12	-.36**	.07	.23	.15	-.02	.06	-.36	.17	.04	.06	-.21	.16
Base - OSPAN	-.19	.17	-.88**	.07	.46*	.15	.04	.06	-.14	.16	.45**	.06	-1.2**	.16
HC - HC + hints	-.14	.17	-13	.07	.18	.15	-.04	.06	.14	.17	.06	.06	-.06	.17
HC - LC	-.08	.17	.06	.07	.03	.15	.04	.06	.29	.17	-.02	.06	.27	.16
HC - LC + hints	.09	.17	-.11	.07	.19	.15	-.04	.06	-.06	.17	.05	.06	.16	.16
HC - OSPAN	-.21	.17	-.64**	.07	.42	.15	.02	.06	.16	.17	.47**	.06	-1.4**	.17
HC + hints - LC	.06	.17	.06	.07	-.15	.15	.08	.06	.15	.17	-.07	.06	.33	.16
HC + hints - LC + hints	.23	.17	.02	.07	.01	.15	.00	.06	-.20	.17	-.00	.06	.22	.16
HC + hints - OSPAN	-.08	.17	-.51**	.07	.24	.14	.06	.06	.02	.17	.41**	.06	-1.4**	.16
LC - LC + hints	.17	.17	.17	.07	.16	.15	-.08	.06	-.36	.17	.07	.06	-.11	.16
LC - OSPAN	-14	.17	-.70**	.07	.39	.14	-.02	.06	-.14	.16	.48**	.06	-1.7	.16
LC + hints - OSPAN	-.12	.17	-.53**	.07	.23	.14	.06	.06	.22	.16	.42**	.06	-1.6**	.16

*MD (mean difference = I-J)- Base: baseline measurement; HC: high load/LC: low load/significant: < 0.01**, < 0.05**

The results for *RQ2* are presented by Table 2. *RQ2a* investigated whether *self-reported cognitive load* across the different sets of exercises can be explained through physiological data. First, in the baseline model, it was observed that the sets with high element interactivity resulted in higher self- reported cognitive load, whereas the provision of hints reduced self- reported cognitive load. Furthermore, in terms of physiological data, results indicated that CL is explained by a significant negative effect of the slope of ST, and a positive effect of mean HR. The total variance explained is r^2=.18 and the model including physiological data was significantly better: χ^2 (7, N=64)=32.19, p<.001 when compared with the baseline model. *RQ2b* investigated whether *task performance* across the four sets of exercises can be explained through physiological data. The baseline model revealed that high task complexity decreased task performance, whereas no effect of the provision of hints was found. In terms of physiological data, results reveal that a negative slope of ST is associated with higher task performance. The total variance explained is r^2=.47 and the model including the physiological data was significantly better: χ^2(7, N=64)=15.31, p<.05.

Table 6.3.

Multilevel analyses with self-reported cognitive load and task performance as dependent variables (*N*=64)

	Self-reported cognitive load		*Task performance*	
	Est. *(SE)*	Est. *(SE)*	Est. *(SE)*	Est. *(SE)*
Intercept	-.19*(.11)*	-.24*(.11)*	.62(.09)	.60*(.08)*
Manipulations				
Complexity (High*)*	.57*(.09)****	.58*(.09)****	-1.34(.08)***	-1.32*(.08)****
Hints	-.21*(.09)**	-.20*(.09)**	.07(.08)	.07*(.08)*
GSR				
GSR_SCmag		.01*(.09)*		.04(.07)
HR and HRV				
ECG_meanHR		.26*(.07)****		-.03*(.06)*
ECG_SDNN		.00*(.06)*		.10*(.05)* .
ECG_LF		.03*(.09)*		-.05*(.07)*
LF_VLF_Ratio		-.11*(.08)* .		.07*(.05)*
ST				
ST_mean		.13*(07)* .		.03*(.06)*
ST_slope		-.14*(.06)**		-.15*(.05)***
Random effects	Variance *(SD)*	Variance *(SD)*	Variance *(SD)*	Variance (*SD*)
Participant	.46*(.68)*	.34*(.59)*	.18*(.43)*	.17*(.42)*
Residual	.50*(.71)*	.50*(.70)*	.37*(.61)*	.37*(.61)*
r^2	.09	.18	.44	.47

*Significant: <0.001***, < 0.01**, < 0.05*

RQ3 aimed at identifying features that assess high (i.e. OSPAN) and low cognitive load (i.e. baseline measurement) by constructing a logistic regression model. Table 4 reveals the importance of the slope of ST and mean HR for distinguishing between low and high cognitive load. Higher values of mean HR and the slope of ST indicate higher (objective) cognitive load. The overall accuracy, sensitivity and specificity were respectively .76, .74 and .79 indicating a good performance of the logistic regression model.

Table 6.4.

The coefficients table of the logistic regression model (high cognitive load; *N* = 124)

	Estimate	*SE*	*P*
(intercept)	.05	.40	.87
GSR			
GSR_SCmag	.01	.19	.53
HR and HRV			
ECG_meanHR	.47	.13	.00***
ECG_SDNN	-.14	.12	.24
ECG_LF	.12	.19	.53
LF_VLF_Ratio	-.15	.11	.20
ST			
ST_mean	.02	.12	.88
ST_slope	.61	.12	.00***

*Significant: <0.001***, < 0.01**, < 0.05**

6.5. Discussion

Main findings and implications

This study aimed at measuring cognitive load through physiological data. Therefore, this study systematically manipulated intrinsic and extraneous load to investigate how physiological features, namely, GSR, ST and HR(V) vary as a result of changes in cognitive load. More particularly, four sets of exercises on probability calculations in statistics were developed (1) high complex set with hints, (2) low complex with hints, (3) high complex without hints and (4) low complex without hints. Additionally, a working memory capacity test (i.e. OSPAN), and baseline measurement (i.e. watching a relaxing movie) was used as verification of resp. high and low cognitive load.

RQ1 investigated whether the different conditions (i.e. sets of exercises, OSPAN and baseline measurement) result in differences in physiological data. Trends in our data revealed that *GSR* increased with increased task difficulty, but no significant differences were observed. Our findings are in line with Larmuseau et al.'s study (2019). Nonetheless, our findings are in contrast with the study of Nourbakhs et al. (2012) indicating that GSR reveal differences in task complexity. These contradictory findings can be explained by the difference in task design between the study of Nourbakhs et al. (2012) and our study. More particularly in Nourbakhs et al.'s (2012) study students were given a working memory task such as OSPAN in our study, in which the difficulty level increased (e.g. the addition of more complex numbers). From the experience of our study, we can assume that cognitive activity does indeed increase as more information needs to be remembered and processed. The advantage of this type of tasks is that the students' prior knowledge is less important, which means that the cognitive load for the students increases more evenly compared to the statistical exercises in our study.

With respect to *HR*, our findings revealed that mean HR was significantly lower during the baseline measurement compared with the other interventions, with the exception of the low complex set of exercises without hints. These results might indicate that providing hints during the low complex set of exercises has provoked additional cognitive load since instructional support might not have been necessary and might even have distracted students (Sweller, 2010). Similarly, results indicated that HR was significantly higher during OSPAN compared to the sets of exercises and the baseline measurement. This might suggest that OSPAN induced more cognitive load than the other interventions. Similarly, Cranford et al. (2014) revealed that HR is higher in case of higher complexity levels. In our study, *HRV* (i.e. SDNN) distinguished the OSPAN and baseline measurement. The fact that a decrease of HRV is associated with high cognitive load was also observed by Taelman et al. (2011) who also used HRV to distinguish between a high complex task (also a working memory task) and a rest phase (also a relaxing movie). However, it seems that HR is a more reliable indicator of task difficulty when compared to HRV, which was also observed in the study of Brouwer et al. (2014).

Regarding *ST*, our findings indicated that mean ST was lower and the slope of ST was higher during OSPAN when compared with the other interventions. However, ST did not vary as a results of the manipulated exercises. This latter finding is in line with the study by Larmuseau et al. (2019) where no differences were detected between a high and low complex task manipulated based on element interactivity. Summarized, against our expectations, the manipulation of the level of complexity of a task, based on two dimensions i.e., intrinsic (i.e. element interactivity) and extraneous load (i.e., provision of hints) did not result in differences in physiological data. As aforementioned, possible explanation for this is that we did not take important student characteristics into account (e.g. prior knowledge). On the other hand, it is also possible that this is due to both the task and study design. For instance, students were allowed to choose whether or not to consult the hints. In addition, the study was also non-committal, which meant that students who found the exercises too difficult did not try to solve them. Accordingly, those students did not experience CL, which biased the results.

Although our findings were not as expected in terms of the developed set of exercises, we observed that the OSPAN induced high CL. In this study, OSPAN was used as a control intervention for continuous (high) CL as this dual-task requires both storage and processing of information (Yuan et al., 2006). Moreover, as all physiological features are also related to stress (see section 2.2), OSPAN might also have induced stress as a reaction on continuous high levels of CL (Herborn et al., 2015; Iani et al., 2004). OSPAN consisted of different trails of remembering and processing information. The more numbers students had to remember and the more mathematical problems they had to solve, the harder it was for students to complete the test correctly which might have induced both high CL and stress. This would be in accordance with the literature since task failure and feelings of lack of control have been shown to induce stress (Conway, Dick, Wang & Chen, 2013).

RQ2 examined to what extent variance in self-reported cognitive load and task performance can be explained through physiological data. In terms of *self-reported cognitive load* (RQ2a), results revealed a positive influence of the level of complexity and a negative influence of the provision of hints which is in line with CLT and indicates that we have succeeded in our manipulations (Sweller, 2010). Results furthermore indicated that a less negative slope of ST is related to the invested cognitive load. Based on the results of RQ1 (i.e. positive slope during OSPAN) this might indicate that high levels of cognitive load might have induced stress. Therefore, we can assume that students who experienced less stress were able to invest more cognitive load. This hypothesis is in line with CLT stating that exceeding the working memory capacity (i.e. cognitive overload) can interfere with learning (Sweller, 2010; van Merriënboer & Sluijsmans, 2009).

Results also revealed that self-reported cognitive load is positively related to higher HR, which is in line with the findings of Cranford et al. (2014). However, our findings also revealed that not much variance in cognitive load was explained by physiological data. This might suggest that self-reported and physiological data provide different information about the learning process. In fact, physiological data relate to reactions that happen largely unconsciously whereas self-reported data are more conscious (Dindar et al, 2019). Nonetheless, the combination of both types of data provides more insight into the meaning of physiological data. In terms of *task performance* (RQ2b), our findings indicated that high levels of complexity negatively influenced task performance. No influence of the provision of support was observed. Again, not much variance in task performance was explained by physiological data. Results showed a negative influence of the slope of ST on task performance. When comparing results of task performance with the results of self-reported CL, we can infer that perhaps less stress led to better task performance. What is striking here is that HR and HRV had no influence on task performance. Presumably, students who were more relaxed during the problem-solving process (i.e. less cognitive load and stress) might also had better knowledge of the content and therefore attained higher task performance.

RQ3 aimed at identifying the most important features for assessing high and low cognitive load. Based on the findings of RQ1 we decided to compare OSPAN (high cognitive load) with the baseline measurement (low cognitive load). Results indicated that higher values in the slope of ST and mean HR are indicative for high cognitive load. These results are in line with findings of *RQ1* and *RQ2,* and further provide evidence that OSPAN required high levels of cognitive load. When we take into account that OSPAN also induced stress, results are in line with the study of Karthikeyan et al. (2012). This study used a working memory test (i.e. Stroop color word test) to investigate whether ST can detect stress. Results revealed that ST was a reliable measure for identifying stress level changes. These findings demonstrate the potential of physiological signals to detect high cognitive load or cognitive overload (i.e. exceeding working memory capacity). Consequently, we have to take into account that in the current study stress might have obscured the relationship between cognitive load and physiological data.

Limitations and suggestions for future research

Regardless of the fact that this study provides insight into the use of physiological data for measuring CL, the study is also characterized by some limitations. A first limitation relates to the fact that there was a lot of missing data which possible affected our results. A second limitation relates to the fact that we did not assess self-reported cognitive load during OSPAN and the baseline measurement. This could have provided a clear insight into learners' perception of cognitive load during these phases and further unravel the association between self-reported and physiological data. A third limitation was the noncommittal nature of the study, and consequently, this might has resulted in students being unmotivated to solve the exercises. Therefore, it might be useful for future research to integrate the study into the regular curriculum. Fourthly, important internal conditions should not be overlooked in order to obtain meaningful information from MMD (Gašević et al., 2015). Students' prior knowledge might have impacted our findings, as domain knowledge (e.g. formula) has a major influence on cognitive load (Sweller et al., 2019). Also affective characteristics, such as negative feelings towards the learning material can induce CL by disrupting working memory processes, particularly for high complex tasks (Basanovic et al., 2018). Therefore, this might have influenced the physiological recordings. A final limitation is related to the machine learning model. The sample size was rather low for using machine learning techniques. Consequently, based on our data we cannot make major statements about the results. Therefore it should be emphasized that it is mainly an indication of feature importance.

6.9. Conclusion

Against our expectations, results revealed that physiological data could not be used to detect differences in cognitive load based on intrinsic and extraneous manipulations. By contrast, most of the significant results are related to OSPAN and the baseline measurement. Based on our findings of OSPAN, we might conclude that HR, HRV and ST is more sensitive to high cognitive load, namely, exceeding the learner's cognitive capacity and the related mental states (i.e. stress). In this respect, as high cognitive load can also provoke stress, it is not always clear what exactly is measured via physiological data. Therefore, it remains important to combine self-reports of associated mental states with physiological data in future studies as this might facilitate interpretation.

Chapter Seven

Discussion and concluding remarks

7. Discussion and concluding remarks

Nowadays, increasingly more courses take place online and incorporate rich authentic whole-tasks that provide opportunities for complex learning (van Merriënboer & Sluijsmans, 2009). A research-based instructional design model that has proven to be effective in promoting complex learning (Lim et al. 2009; Melo & Mirando; Sarfo & Elen, 2005), and can be integrated into online learning environments (Frerejean et al., 2019a; Melo & Miranda, 2014), is the 4C/ID-model (van Merriënboer et al., 2002). Nonetheless, offering an online learning environment based on the 4C/ID-model is no guarantee for its effectiveness (Elen, 2020). As the learner is perceived as an active agent in accomplishing learning processes, the effectiveness of online learning environments largely depends on students' cognitive and motivational-affective characteristics (e.g. Elen, 2020; Jiang et al., 2010; Martens et al., 2004). In order to investigate factors that can improve the effectiveness of online courses, the current PhD project was divided into respectively research track 1 and 2.

Research track 1 investigated the influence of students' cognitive and motivational-affective characteristics on the effectiveness of a 4C/ID-based online course. Former research indicated that the effect of individual differences can be monitored by aligning the learning environment with students' learning needs (Clarebout & Elen, 2007; Moos & Azevedo, 2009). In view of aligning the online course with students' cognitive needs, and given the focus on complex learning, *research track 2* explored methods for investigating and measuring cognitive load during the problem-solving process by using multimodal data (including physiological data).

As the research project is divided into two research tracks that have largely different research aims, these research tracks will first be discussed separately. In the first part, the main findings of the three studies from research track 1 are explained, combined with the limitations and directions for future research and, theoretical and practical implications. In the second part, research track 2 is approached similarly. Finally, general conclusions are formulated across the two research tracks.

Research Track 1: individual differences determining the effectiveness of 4C/ID-based online courses

The research aim of research track 1 was to investigate the influence of students' cognitive and motivational-affective characteristics on students' use of 4C/ID-based online courses and subsequently their learning outcomes. As such, this section will first elaborate on the impact of students' cognitive characteristics on the use of the learning environment, next on the influence of students' motivational-affective characteristics on that use, and finally the influence of use on students' learning outcomes.

7.1. Main findings of Research Track 1

7.1.1. Influence of cognitive characteristics on use

In the studies of research track 1, two online courses were systematically designed according to the 4C/ID-model in view of promoting complex learning (van Merriënboer et al., 2002; van Merriënboer & Kirschner, 2018). However, as the learners are active agents in their learning process, the effectiveness of online courses largely depends on the learners' characteristics (Liem et al, 2008; Martens et al., 2007). Students' characteristics are even more important in an online context where a lot of autonomy is required from the learners (Tsai, 2013). Autonomy in online courses can largely be manipulated by the amount of *learner control* that is provided (Väljataga & Laanpere, 2010). In the two online courses that were developed for the studies in research track 1, the components of the 4C/ID-model were offered in a non-embedded manner. Consequently, a large amount of learner control was provided which enabled learners to self-direct their learning. More particularly, students were able to choose their own learning trajectory and select components in line with their personal preferences and cognitive needs (Opfermann, Scheiter, Gerjets, & Schmeck, 2013; van Merriënboer & Sluijsmans, 2009).

In order to investigate whether students selected the components in line with their cognitive needs, Study 3 investigated the influence of *students' prior knowledge* on the *differences in use* of the four components of a 4C/ID-based online course. Results revealed that there was a negative relationship between students' prior knowledge and part-task practice. More particularly, the lower students' prior knowledge was, the more students consulted part-task practice. *Part-task practice* provided opportunities for additional practice of routine subskills of the learning tasks. As a result, these findings seem to suggest that learners with lower prior knowledge realized they did not have the required level of routine subskills (van Merriënboer et al., 2002). By contrast, students with lower prior knowledge did not significantly make more use of supportive information, although this was advisable according to the 4C/ID-guidelines. More particularly, *supportive information* allows for constructing cognitive schemas in long-term memory which may be absent from students with lower domain-specific prior knowledge (van Merriënboer & Kirschner, 2018). Therefore, findings indicate that differences in students' cognitive characteristics impact a different use of the components of the 4C/ID-model. Whereas, the larger use of part-task practice is in line with our expectations (i.e. based on the guidelines of the 4C/ID-model), this is not the case for the use of supportive information. When students do not engage as expected and/or do not adequately use the provided support, this phenomenon can be perceived by the instructional designer as *'instructional disobedience'* (Elen, 2020). Although we do not claim that students in study 3 were 'instructional disobedient', our findings suggest that students chose to consult part-task-practice to remediate their knowledge gaps instead of supportive information. Possibly, students felt they mainly lacked the routine subskills of the learning tasks (i.e. covered in part-task practice) instead of the non-routine subskills (i.e. covered in supportive information) (van Merriënboer & Kirschner, 2018).

7.1.2. Influence of motivational-affective characteristics on use

Research track 1 also investigated the influence students' motivational-affective characteristics on the quantity and quality of use (i.e. Study 1,2) and differences in use (i.e. Study 3). Based on the Expectancy-Value Theory, self-efficacy and task value are assumed to influence academic engagement (Wigfield & Eccles, 2002). As such, in Study 3, the influence of *task value and self-efficacy* on *differences in use* of the four components of the 4C/ID-model was investigated. In terms of task value, results revealed a positive relationship between students' task value and the consultation of learning tasks and supportive information. In Study 3, *task value* identified students' interest in the subject matter and how much the student valued the desired outcome (Duncan & McKeachie, 2005). More particularly, this implies that students that were intrinsically motivated put more effort into solving the learning tasks (Chen & Jang, 2010). Moreover, students' task value influenced the consultation of supportive information. As supportive information is known for connecting present knowledge with novel knowledge, this seems to imply that students with higher task value were more eager to learn new things. This finding is also in line with previous studies showing that students with greater interest, consciously explore additional learning materials (Bong, 2001; Martens et al., 2004). In terms of self-efficacy, despite our expectations, we did not observe an influence of self-efficacy on differences in the use of the four components. Self-efficacious students believe they are capable of executing the learning tasks (Zimmerman, 2000). Consequently, although self-efficacy is believed to be an important characteristic for influencing students' engagement we did not find empirical evidence for this assumption (Taipjutorus et al., 2012). Nevertheless, it is possible that the effect of self-efficacy in Study 3, might have been outweighed in the research model by students' prior knowledge as these constructs were moderate related (Kline, 2013).

Findings of Study 3 emphasize the importance of students' prior knowledge and motivation for grasping learning opportunities. In addition, former research also emphasized the importance of technology acceptance (Šumak et al., 2011). In order to retrieve information on the *technology acceptance* of the 4C/ID-based online course, PU and PEOU were adopted from TAM (Davis, 1989) in Study 1 and 2. According to TAM, PU is defined as the extent to which learners believe that the online course is a useful learning tool for enhancing their performance. Whereas, PEOU is defined as the extent to which learners believe that using the online course does not entail extra mental effort. Study 1 examined how technology acceptance influenced students' quantity of use (i.e. total time spent) of the 4C/ID-based online learning environment for teaching French as a foreign language. The study was carried out within the teacher training program for primary school education in which learning how to teach French was part of the students' training program. Findings suggested that students' PU of the online learning environment is related to the quantity of use (i.e. time spent). The results correspond with a former study of Juarez Collazo et al. (2012) that also indicated that PU positively influences the use of the online learning environment. Regarding PEOU, findings of Study 1 indicated that PEOU did not influence the actual use of the online course. This might indicate that user-friendliness is subordinate to perceiving the online learning environment as a useful learning tool for learning the subject matter.

The second study investigated the relationship between *technology acceptance* and the *quantity of use* (i.e. total amount of course activity) and *quality of use* (i.e. course performance) of an online learning environment for learning French as a foreign language. The results of Study 2 revealed that there was no significant relationship between both PU and PEOU on the quantity and quality of use. The mixed findings of Study 1 and 2 might be related to the fact that the subject matter in Study 2 was not integrated into the students' training program. These findings emphasize the importance of authentic learning designs (i.e. social and physical context in which it will be used). Former research and educational theory already indicated that authentic learning designs have the potential to improve students' engagement (Brown, Collins & Duguid, 1989; Herrington et al., 2003; Martens et al., 2004; van

Merriënboer & Kirschner, 2018). So despite the use of authentic tasks, learning French may have been less relevant to the students' everyday lives in Study 2 when compared with Study 1.

In the second study, TAM was expanded with an external variable, namely, *perceived instructional quality*. In order to retrieve information on students' perceived instructional quality, students were asked to fill in the Teaching and Learning Questionnaire (TALQ) developed by Frick et al. (2009). This questionnaire incorporates the five *First Principles of Instruction* of Merrill (2002). Merrill's (2002) First Principles of Instruction define that learning is promoted when learners are engaged in solving real-world problems, existing knowledge is activated, new knowledge is demonstrated, applied, and integrated into the learners' world. Findings of Study 2 indicated that the *perceived instructional quality* had a significant positive influence on students' PU and PEOU (i.e. technology acceptance). These findings seem to indicate that technology acceptance and the perceived instructional quality of the online course are largely interrelated. This relates to former studies that plea for the extension of the TAM model with pedagogical quality (e.g. Lee et al., 2009; Liaw & Huang, 2013; Yang et al., 2017).

Additionally, Study 2 investigated the relationship of *technology acceptance* and *perceived instructional quality* on the *quality of use* (i.e. course performance). Findings indicated no association between technology acceptance and course performance, whereas students' perceived instructional quality was related to course performance. This latter finding indicates that learners who recognized the pedagogical quality invested more mental effort into solving the exercises qualitatively. Findings can be related to the study of Frick et al. (2010) who indicated that integrating the First Principles of Instruction of Merrill (2002) in the course design is related to higher course performance. However, it is not excluded that students who had a greater knowledge of French, also had a better understanding of the relevance of the learning activities, and therefore were more motivated to attain pre-established goals (Elen, 2020).

7.1.3. Influence of use on learning outcomes

In Study 1 and 3 the influence of respectively the *quantity of use* and *differences in use* of the 4C/ID-based learning environment on students' learning outcomes was investigated. Findings of Study 1 indicated that the *quantity of use* of the online course, influenced students' learning outcomes when controlling for students' prior knowledge. Additionally, Study 3 indicated that *differences in use* also improved students' learning outcomes, when controlling for students' cognitive and motivational characteristics. More specifically, the combination of the *activity on learning tasks* and the *consultation* *of procedural information* seemed to improve students' learning outcomes. The procedural information provided instructional guidance which helped learners to perform the routine subskills of learning tasks. Procedural information was offered just-in-time while solving the learning task and as such improved learning (van Merriënboer & Kirschner, 2018). Based on the Cognitive Load Theory (CLT) this might indicate that the provision of instructional guidance prevented learners from focusing on irrelevant aspects. As such, the students were able to devote more working memory resources to learning (Sweller, 2010; van Merriënboer & Kirschner, 2018). Nevertheless, we must also point out that given the high relationship between prior knowledge and students' learning outcomes, there must have been little learning gain. This might indicate that students mainly solved learning tasks they already mastered. This may also partly explain why there is no relationship between supportive information (i.e. includes new subject matter), part-task practice (i.e. practicing routine subskills) and learning outcomes. Students probably invested little effort in learning and practicing new subject matter. This might have been the result of the fact that the subject matter of Study 3 was not part of the students' training program (Brown et al., 1989).

An overview of the findings are illustrated in *Figure 1*. The arrows indicate all the relationships that were examined across research track 1. The solid line arrows indicate significant relationships, whereas the dotted line indicate the non-significant.

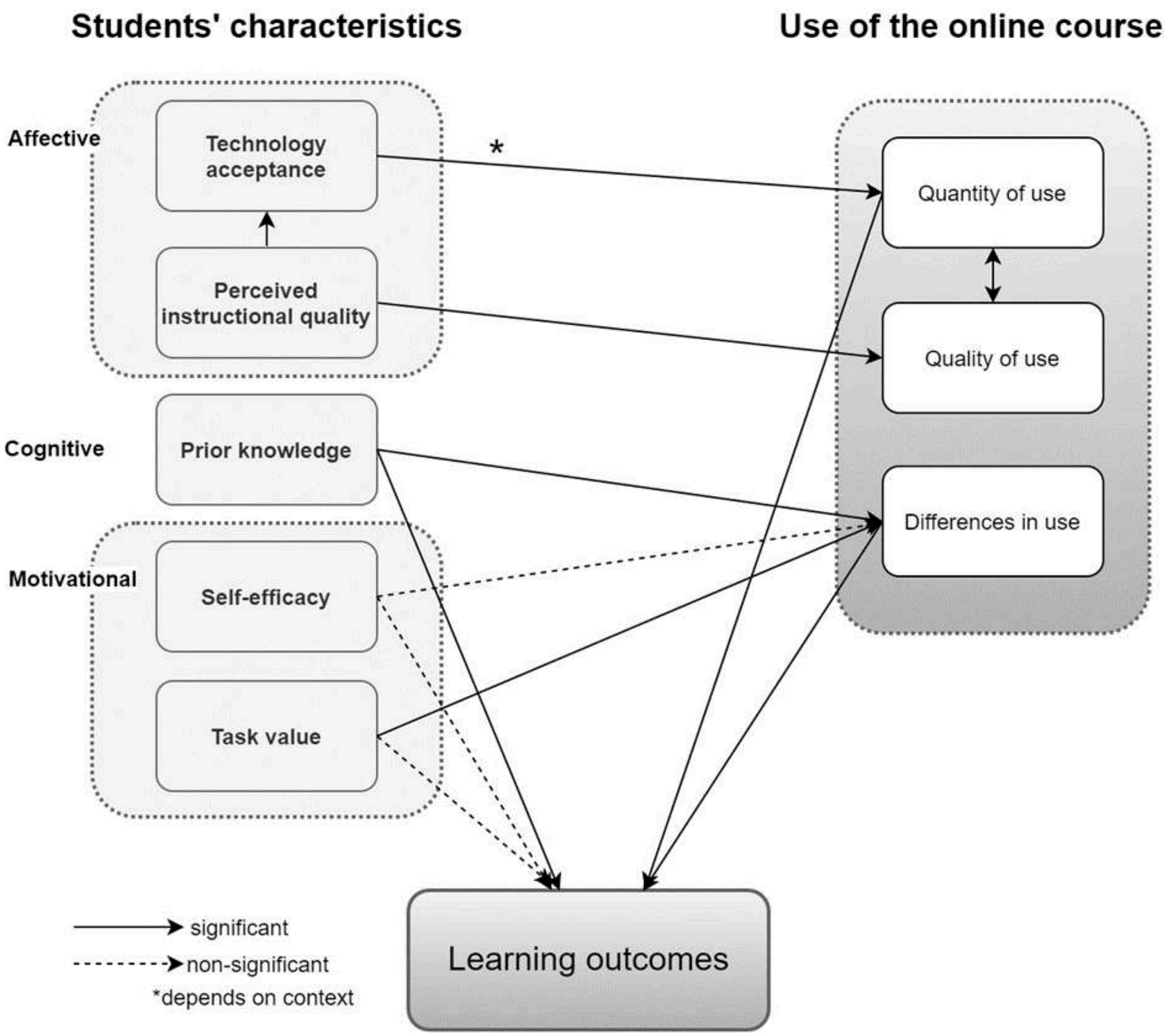


Figure 7.1.: Overview of the Findings of Research Track 1

7.2. Limitations and future directions

A major *first limitation* across the three different studies in research track 1, is that information on students' learning outcomes was retrieved from paper and pencil pre-and posttests (Evens et al., 2017). These tests mainly concerned knowledge vocabulary and grammar. Consequently, these tests do not match the definition of complex learning defined by van Merriënboer et al. (2002). More particularly, important skills (e.g. listening skills) and attitudes (e.g. forms of politeness in French) that were covered in the online course were not surveyed in an integrated manner. Additionally, the studies did not investigate whether students were able to transfer what they have learned into daily life or work situations. As such, the findings of research track 1, only provide limited information about whether or not complex learning has been achieved. Future research should retrieve information on complex learning by evaluating students' knowledge, skills, and attitudes during real-life practice.

A *second limitation* relates to the collection of data via the Moodle LMS. Specifically, the first two studies applied two different manners to determine *the quantity of use*. In Study 1, the quantity of use was the total time spent. Nonetheless, *time spent* can be biased as the retrieved time spent by the LMS is often different than the actual time spent (Frerejean, Velthorst, van Strien, Kirschner, & Brand-Gruwel, 2019b). For instance, time spent was biased by students who did not close their browsers when they were no longer using the online learning environment. To reduce this bias, time spent was checked by manually parsing trough the data and checking the length of time between the logged activities. When the time spent between two activities was unrealistic (i.e. when compared with other students), data was deleted. Consequently, collecting *time spent* in Study 1 required subjective judgment. In addition, it is also a disadvantage that the LMS registers activities per minute. This might imply that the reading time of procedural information was not accurately recorded as reading procedural information might take less than a minute. Due to the limitations of using time spent as an indication of the quantity of use, Study 2 and Study 3 recorded the total number of statements (i.e. *course activity*) to define the quantity of use. These statements include quiz attempts, submissions, consultation of pages, etc.

A *final limitation* is that no different versions of 4C/ID-based online courses were compared. This prevents us from making recommendations regarding the instructional design. Follow-up studies should compare different versions of 4C/ID-based online courses. For instance, it might be interesting to compare the effectiveness of online courses with a different amount of embedded and/or non-embedded support. From this research, it would be more clear if embedding the four components in an online learning environment would lead to higher performance.

7.3. Implications of research track 1

In general, findings of Research Track 1 reveal that individual differences influence the effectiveness of a 4C/ID-based online learning environment for complex learning. In the context of information-processing technology (IPT), Perkins already indicated in 1985 that learning opportunities are taken when three conditions are met. These conditions are (1) the provision of the learning opportunity, (2) the recognition of the learning opportunity and, (3) motivation to take the learning opportunity. The first condition mainly relates to the physical presence of the learning opportunity, while the other two conditions are more related to students' motivational-affective characteristics (Elen, 2020). In the current research project, the provision of the 4C/ID-based online course can be perceived as the learning opportunity. In line with Perkins (1985), findings of research track 1 also seem to emphasize the importance of technology acceptance and the perceived instructional quality of the online learning environment for complex learning. Students' perceptions towards technology acceptance can be perceived as *technological functionality* whereas, perceived instructional quality can be perceived as *pedagogical functionality*. More particularly, higher levels of students' perceived technological and pedagogical functionally seem to influence the use of the online course in a positive manner. Additionally, findings also illustrated that students use the online course differently, based on their cognitive and motivational needs. It seems that mainly students' *motivation* is important when it comes to learning new things. This is in line with the third condition for grasping learning opportunities according to Perkins (1985). Overall, findings of research track 1 are very much in line with the suggested conditions of Perkins (1985) for grasping learning opportunities. Knowing the importance of these learner attributes can assist instructional designers and/or lecturers in designing effective online learning environments that meet students' needs.

Research track 1 also emphasizes the need for multimodal data when investigating students' online learning processes. Research track 1 reveals that LMS *log-data* (i.e. records of users' activity) can easily be captured unobtrusively. Moreover, collecting LMS log-data provides insight into the use of the online course. Nonetheless, it should be noted that former research indicated that log-data is not a significant proxy measure of the invested mental effort (Henrie, Bodily, Larsen & Graham, 2018). For instance, a given activity might require more mental effort (and subsequently more time) from a student with lower prior knowledge (Moissa et al., 2019). As such, merely relying on log-data and not taking *the internal and external conditions* into account to investigate the exerted mental effort, might provide biased information (Gašević et al., 2015). Findings of research track 1 illustrated that the relationships between students' characteristics and log-data provide more reliable insight into the use of the online learning environment. Additionally, when investigating the influence of students' characteristics on the different components, this provided information on students' self-directed learning in accordance with their learning needs. However, findings also reveal that additional information (that is collected during the online learning process) is necessary to have insight into the extent to which students are conscious regarding their self-directed learning. For instance, it is not clear from the findings whether students consulted support when they perceived high cognitive load. As a result, collecting real-time multimodal data would help us to unravel the *black box* of the learning processes.

Research track 2: MMLA for measuring cognitive load during online complex problem-solving

The research aim of research track 2 was to investigate and measure cognitive load by using multimodal data. In this section we will first elaborate on the link between research track 1 and 2. Secondly, this section discusses the main findings. More particularly, we will discuss the relationship between self-reported data and physiological data for assessing cognitive load. Subsequently, the usability of physiological data for assessing cognitive load will be discussed.

7.4. Transition of research track 1 to research track 2

Findings of research track 1 indicated that the effectiveness of online courses for complex learning is largely influenced by students' individual differences. It could therefore be an added value to make the learning environment more responsive to the learning needs of the students and subsequently, improve students' learning outcomes. Student modeling is a proven technique to support individual differences (Yelizarov & Gamayunov, 2014). Based on the student model, *intelligent tutoring systems* (ITS) can be built that monitor students' learning processes as illustrated in *Figure 2*. The *student model* is the core model that should collect information on students acquired from trace data. The *domain model* is the source of expert knowledge that can be used to evaluate students' cognitive states, task performance etc. The *tutoring model* receives information from the domain and student model, and should make decisions on whether or not to intervene and how (Nkambou, Bourdeau, & Mizoguchi, 2010). Given the demands that complex learning puts on students' working memory, it might be advisable to develop an *interactive online learning environment* based on students' *cognitive needs*. For instance, students with lower prior knowledge that experience high cognitive load might benefit from customized support such as worked examples or tasks with lower element interactivity etc. (Chen & Kalyuga, 2020; Sweller, 2010). In contrast, students with higher prior knowledge might benefit from the omission of support and a faster transition to more difficult tasks (van Merriënboer & Sluijsmans, 2009). In order to be able to design a complex interactive online learning environment we should be able to find reliable indicators of the perceived cognitive load that can be measured in real-time and in a non-obtrusive manner (Frerejean et al., 2019b; Moissa et al., 2019).

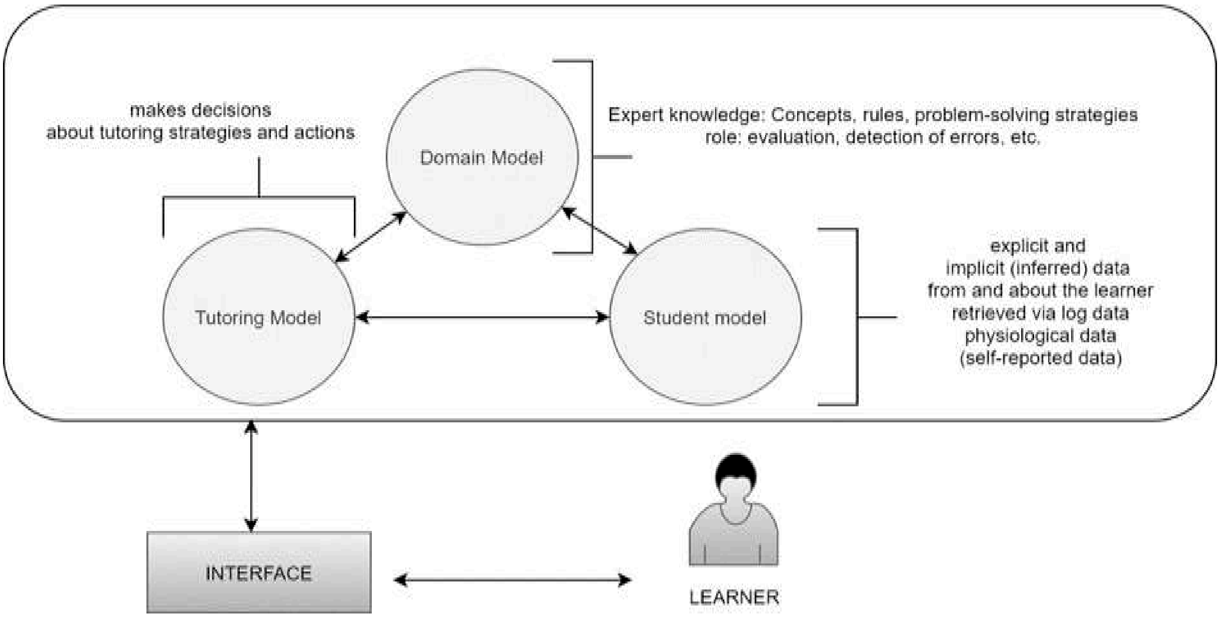


Figure 7.2: The four-component architecture of an ITS (Nkambou et al., 2010, p.5)

Former researchers already showed interest in using high-frequency sensing technologies to assess real-time cognitive load in a less invasive manner (Chen et al., 2016; Worsley, 2018). More particularly, several studies have indicated that *physiological measures* can detect cognitive load. Physiological approaches for assessing cognitive load are based on the assumption that changes in cognitive load can be reflected by body properties. Several studies indicated that an increase in cognitive load leads to increases in *electrodermal activity* (EDA), decreases in *skin temperature* (ST), increases in *heart rate (*HR) and decreases in *heart rate variability* (HRV; e.g. Haapalainen et al., 2010; Or & Duffy, 2007; Shi et al., 2007; Taelman et al., 2011). Moreover, the use of wearable devices allows us to assess these skin-response measures and cardiovascular measures in a non-obtrusive manner and in real-life situations (Moissa et al., 2019). As a result, research track 2 focused on investigating whether cognitive load can be measured by using physiological data.

7.5. Main findings of Research Track 2

Research track 2 explored methods on how to investigate and measure cognitive load during online complex learning. To induce cognitive load, two studies experimentally manipulated intrinsic and/or extraneous load. In Study 4, the *intrinsic load* was manipulated by changing task complexity based on the extent of element interactivity which resulted in a high and low complex task. In Study 5, both the *intrinsic and extraneous load* was induced by respectively manipulating the level of complexity and the presence or absence of procedural information. The manipulations were respectively: low cognitive load with procedural support (i.e. step-by-step instruction), high cognitive load with procedural support, low cognitive load without procedural support and, high cognitive load without procedural support. During the interventions, students were allowed to consult an online version of their manual. Both studies had a within-subject design in which each participant participated in the conditions in a randomized order. In view of measuring and investigating cognitive load during online complex problem solving, both studies combined physiological data with self-reported cognitive load.

7.5.1. Self-reported cognitive load versus physiological data

A traditional and simple way to measure cognitive load is via subjective self-reported data. Subjective data provides information about the learner's perceived cognitive load during the learning activity (Paas, 1992). Nonetheless, these perceptions do not always match the actual cognitive load (Boekaerts, 2017; Dindar et al., 2019; Vanneste et al., 2020). To investigate to what extent self-reported data matches physiological data for assessing cognitive load, the two studies in research track 2 combined self-reported data with physiological data retrieved form wrist-worn wearables measuring EDA and ST (Study 4) and chest-patches measuring HR and HRV (Study 5).

More specifically, in Study 4, the self-reported cognitive load was measured by using a multidimensional subjective cognitive load rating scale developed by Leppink et al. (2013) for the distinct types of cognitive load, namely intrinsic, extraneous, and germane load. Additionally, to have insight into the total amount of exerted mental effort (i.e. the overall perceived cognitive load), the unidimensional mental effort rating scale developed by Paas (1992) was used. Since the exercises in Study 4 were manipulated on the basis of *element interactivity*, we first verified this manipulation. Findings revealed that we succeeded in manipulating the intrinsic load. Additionally, a bivariate correlation analysis was conducted between the perceived *intrinsic load* and physiological data (i.e. EDA and ST) during a high and low complex task (Sweller, 2010). Findings of Study 4 revealed no significant correlations between the perceived intrinsic load and physiological data. This might be due to the fact that the psychometric scale developed by Leppink et al. (2013) mainly provides insight into cognitive load retrieved from the instructional design (i.e. task complexity, clear instructions etc.). This implies that the allocated working memory resources, influenced by factors beyond instructional design decisions, such as students' prior knowledge, motivation etc., are overlooked. For example, it is possible that a student may have described an exercise as complex, but due to lack of motivation has invested little mental effort in solving that exercise. Therefore, physiological data (i.e. ST and EDA) were also correlated with self-reported mental effort and the overall cognitive load. Findings revealed that there was a high significant correlation between EDA and the overall cognitive load. Nonetheless, this was only the case during the highly complex task and not during the low complex task, which might be an indication that EDA is only sensitive to experiencing a high cognitive load.

Study 5 also investigated the relationship between self-reported cognitive load and physiological data (i.e. EDA, ST, HR, HRV). In Study 5, a multilevel analysis was conducted in order to investigate how much variance of the self-reported cognitive load was explained by physiological data. Results revealed that a significant amount of variance of the self-reported cognitive load was explained by mean HR and the slope of ST. Nonetheless, the total variance explained was very low indicating the limited usability of HR and ST as measurements for the perceived cognitive load. With regard to HRV and EDA, no significant results were found. The contradictory findings between Study 4 and 5 may be largely related to differences in the task and study design. The main difference is that the different levels of complexity between the high and low complex tasks were more pronounced in Study 4 when compared with Study 5 (i.e. larger differences in element interactivity). In addition, in study 5 the students were asked to solve as many exercises as possible in 10 minutes which may have caused stress.

Based on these findings, it is recommended not to make any conclusive statements about the usefulness of EDA for measuring cognitive load. In line with the findings of Study 4, there are different studies that confirmed the interrelationship of EDA and self-reported data. More particularly, a recent study by Vanneste et al. (2020) induced three different levels of cognitive load based on element interactivity (i.e. low, high, and overload condition). By conducting multilevel analyses, the authors investigated how much variance is explained by physiological data (i.e. including ST and EDA). Findings revealed that a significant amount of variance of the self-reported cognitive load was explained by EDA. In addition, a recent study by Johannessen et al. (2020) investigated correlations between self-reported cognitive load and different physiological signals (i.e. including EDA and HRV) of five physicians during real-life resuscitation. Findings revealed that overall, EDA seemed to have the strongest correlation with self-reported cognitive load measured by the Paas' mental effort rating scale (1992). As a result, it seems possible that the differences in cognitive load were too small and that the assessment of cognitive load via EDA in Study 5 was influenced by other emotions (e.g. stress) that occurred during the study (see also section 10.1.2.).

7.5.2. Physiological data for assessing cognitive load

In Study 4, we investigated whether the manipulation of the level of complexity of a task (i.e. high and low complex task) resulted in differences in physiological data (i.e. EDA and ST) when controlled for students' prior knowledge. Findings highlight the potential of EDA for measuring differences between no (or very low) cognitive load (i.e. retrieved from the baseline measurement) and the manipulations of cognitive load. Therefore, this could be a confirmation that EDA primarily detects large differences between cognitive load. Additionally, EDA peaks seem to be related to specific events such as processing instructions, consulting supportive information, etc. In addition to EDA and ST, Study 5 also investigated differences of HR and HRV across the different interventions. Results revealed no clear differences in all physiological data across the four interventions. As far as EDA was concerned, descriptive statistics appeared to be in line with the proposed manipulations (i.e. EDA was higher during high cognitive load) but the results were not significant. In line with previous findings, results seem to indicate that small differences in complexity are not measurable on the basis of EDA. Moreover, it seems that these small differences can also not be assessed by HR, HRV, and ST. Another possibility is that the students might have experienced stress during the study (e.g. because of the time restriction). A prior study conducted by Chen et al. (2016) investigated the effect of stress on cognitive load measurements using EDA as a physiological index of cognitive load. The study experimentally incorporated feelings of lack of control, task failure and, social-evaluation to induce stress. Results revealed that elevated cognitive load resulted in an increase in EDA. Nonetheless, Chen et al. (2016) indicated that the relationship was confounded when the subjects experienced fluctuating levels of stress. As a result, the study of Chen et al. (2016) also seems to suggest that feelings of stress can disrupt the measurement of cognitive load via physiological data. In addition to current findings, it illustrates the sensitivity of physiological data to other emotions.

In Study 5, physiological data were also collected during a computer-based operation span test (OSPAN) in addition to the four conditions. OSPAN was added in this study as a *verification of very high cognitive* *load and/or cognitive overload*. More particularly, OSPAN involves retaining an increasing amount of elements (i.e. numbers), while processing mathematical operations. Consequently, OSPAN was assumed to be effortful and demanding of the working memory (Yuan et al., 2006). Findings of Study 5 revealed that there were significant differences in physiological data visible between the manipulated conditions and OSPAN. Respectively, ST was significantly lower and HR was significantly higher during OSPAN when compared with the four conditions. The bulk of former studies related these ST and HR to stress (e.g. Herborn et al., 2015; Smets et al., 2018). Consequently, it is possible that due to its high cognitive demanding characteristics, OSPAN induced feelings of stress (Zhou, Yu, Wang & Arshad, 2018). Some essential differences between the task design of the four interventions and OSPAN might explain these findings. For instance, during the four conditions, students had a lot of opportunities to manage their cognitive load. More particularly, they could select which exercise they wanted to complete from a series of exercises, and they had the freedom to consult an online manual. Moreover, some students just quit when they perceived the exercises as too difficult. In contrast, OSPAN continued automatically and increased in difficulty. Consequently, all students were forced to continuously perform tasks that required a lot of working memory resources. As a result, students did not have opportunities to manage their cognitive load during OSPAN, which might have resulted in working memory resources depletion (Chen & Kalyuga, 2020).

In view of investigating the feasibility of detecting cognitive load during the complex online problem-solving process, all physiological features (i.e. skin response and cardiovascular measures) were combined. Former studies indicated that the combination of different features greatly improves the performance of the machine learning models for cognitive load detection (e.g. Herbig et al., 2020). Therefore, in Study 5, a machine learning model was applied to conduct a binary classification consisting of all physiological features collected during the baseline measurement (i.e. no or low cognitive load) and OSPAN (i.e. very high cognitive load and/or overload). Results revealed a machine learning model of good performance that indicated that ST and HR were the most important features (i.e. high absolute value of the feature weight) for detecting very high cognitive load and/or overload. Note, that this again might indicate that cognitive overload induces feelings such as stress.

7.6. Limitations and future directions

The first and main weakness of both studies in research track 2, is the low sample size. This ensures that we have to be very cautious when making statements about the results. Moreover, the small sample sizes also prevented the use of advanced computational techniques for detecting cognitive load. Larger datasets would provide more opportunities to build various and more complex machine learning models of which we can compare the performance (e.g. Smets et al., 2018b). In addition, a larger sample size would also allow us to use a higher number of classifications such as low cognitive load, high cognitive load, and cognitive overload, whereas we restricted our analyses in terms of low and very high cognitive load.

A second weakness is that log-data (i.e. task performance and course activity) was not included in the studies in research track 2. It might be interesting for follow-up studies to include log-data into predictive models. More particularly, task performance and course activity can also provide information about the perceived cognitive load. For instance, a study by Ahmad, Robb, Keller, and Lohan (2020) indicated that a decrease in the participants' task performance was related to an increase in cognitive load. Additionally, in Study 5, it was observed that a decrease in students' course activity might relate to an increase in cognitive load and/or overload. For instance, some students who perceived the learning material as too difficult just quit. As such, log-data might be an important feature to optimize the performance of predictive models.

7.7. Implications of research track 2

In summary, the results seem to imply that EDA is related to self-reported cognitive load. Nonetheless, results indicate that small differences in cognitive load are not measurable via EDA. Additionally, cognitive overload seems to induce stress which is detectable by ST and HR.

This conclusion can have some important *implications for the assessment of cognitive load*. Firstly, results emphasize that cognitive load remains difficult to assess due to its highly-dynamic and multidimensional nature (Chen et al., 2016). An attempt was made in Study 4 to capture the multidimensionality of cognitive load by using psychometric scales that measure the three distinct load types (Leppink et al., 2013). Nonetheless, findings revealed that the multidimensional scale of Leppink et al. (2013) largely overlooks the fact that cognitive load is also influenced by factors beyond the instructional design. More precisely, the perceived cognitive load is the result of the interrelationship between the learners' internal conditions and external conditions as illustrated in *Figure 3*. As a consequence, findings seem to indicate that it makes more sense to use the self-reported mental effort scale (Paas, 1992) when capturing the total amount of allocated working memory resources (i.e. cognitive load). Certainly, when compared to scales that only provide detailed information on different aspect of the instructional design (i.e. task complexity, clear instructions etc.).

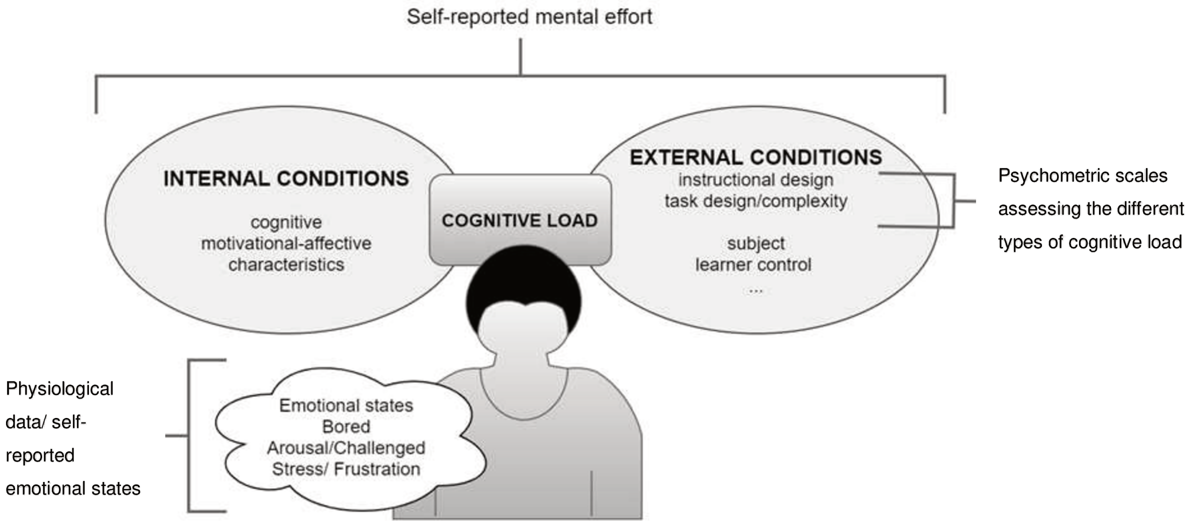


Figure 7.3.: Illustration of the complexity of cognitive load

Additionally, the dimensionality of cognitive load also has *implications for assessing cognitive load via physiological data*. Findings of research track 2 indicate that physiological data is very sensitive to emotional changes that are directly or indirectly related to the perceived cognitive load. More particularly, findings of research track 2 indicated that an increase in cognitive load was detectable by EDA, whereas cognitive overload (which probably induced stress) was detectable by ST and HR (Johannessen et al., 2020; Smets et al. 2018a). Consequently, findings emphasize the need for triangulating EDA analyses with other data sources to provide more robust presentations of cognitive load (Chen et al., 2016).

In sum: let's keep it complex

Previous studies have indicated that the 4C/ID-model is an effective model for designing learning environments for complex learning (Lim et al., 2009; Melo & Miranda, 2015; Sarfo & Elen, 2005). Nonetheless, the provision of qualitative research-based design by no means ensures its effectiveness (Elen, 2020). As learners are active agents in their learning processes, the effectiveness is largely related to students' characteristics (Noroozi et al., 2020). Findings of research track 1 demonstrate, among others, the importance of perceived pedagogical and technological functionality in view of understanding students' quantity and quality of use of complex learning environments. Additionally, findings indicated that the quantity and quality of use influence students' learning outcomes. This demonstrates the importance of a research-based quality instructional design based on authentic learning tasks. In addition, it also shows that it is important that students find the online learning environment a meaningful learning tool to teach the complex subject matter. Additionally, findings of research track 1 indicate that students' prior knowledge and motivation can influence differences in use and emphasize the importance of students' motivation for complex learning. In sum, findings of research track 1 indicate that the effectiveness of online courses for complex learning is largely influenced by individual differences. This might indicate that students would benefit from a more complex online interactive learning environment that monitors students' learning processes.

In order to align the online learning environment with students' learning processes, research track 2 explored methods to measure cognitive load during the online problem-solving process. Results revealed that EDA might be sensitive to large differences in cognitive load. Additionally, results indicated that cognitive overload can result in feelings of stress which is detectable by ST and HR. Therefore, physiological data seems to be a promising avenue to detect real-time cognitive load. Nonetheless, findings of research track 2 also indicate that detecting cognitive load via physiological data is highly complex due to the multidimensionality of cognitive load and the sensitivity of physiological to other emotions and/or mental states. As such, measuring cognitive load seems to remain the holy grail of CLT. Follow-up research needs to clarify which physiological data can be used to detect cognitive load. In addition, it might also be useful for future research to combine cognitive, motivational (e.g. engagement) and emotional learning processes (e.g. stress, feelings of anxiety etc.) as the quality of learning depends on the complex relationships between these learning processes and the external conditions (e.g. online learning; Gašević, et al., 2015; Noroozi et al., 2020). As such, in a distant future, complex interactive learning systems can be designed that meet the students' various learning needs.

References

Ahmad M.I., Robb D.A., Keller I., Lohan K. (2020). *Towards a Multimodal Measure for Physiological Behaviours to Estimate Cognitive Load*. In: Harris D., Li WC. (eds) Engineering Psychology and Cognitive Ergonomics. Mental Workload, Human Physiology, and Human Energy. HCII 2020. Lecture Notes in Computer Science, vol 12186. Springer, Cham. https://doi.org/10.1007/978-3-030-49044-7_1

Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010). Using Electroencephalography to Measure Cognitive Load. *Educational Psychology Review, 22,* 425-438. https://doi.org/10.1007/s10648-010-9130-y

Antonietti, A., Colombo, B., & Di Nuzzo, C. (2015). Metacognition in self-regulated multimedia learning: integrating behavioural, psychophysiological and introspective measures. *Learning, Media and Technology, 40*, 187-209. https://doi.org/10.1080/17439884.2014.933112

Azur, M.J., Stuart, E.A., Frangakis, C., & Leaf, P.J. (2011). Multiple imputation by chained equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research.* https://doi.org/10.1002/mpr.329

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390-412. https://doi.org/10.1016/j.jml.2007.12.005

Baker, R. S., & Inventado, P. S. (2014). *Educational data mining and learning analytics.* In Learning Analytics: From Research to Practice. https://doi.org/10.1007/978-1-4614-3305-7_4

Barnard-Brak, L., Paton, V. O., & Lan, W. Y. (2010). Self-regulation across time of first-generation online learners. *ALT-J: Research in Learning Technology.* 61-70. https://doi.org/10.1080/09687761003657572

Basanovic, J., Notebaert L., Clarke P.J.F., MacLeod C., Jawinski P., Chen N.T.M. (2018) Inhibitory attentional control in anxiety: manipulating cognitive load in an antisaccade task. *PLoS ONE, 13.* https://doi.org/10.1371/journal.pone.0205720

Bell, B. S., & Federman, J. E. (2013). E-learning in postsecondary education. *The Future of Children, 23,* 165-185. https://muse.jhu.edu/article/508225

Benbasat, I., & Barki, H. (2007). *Quo vadis TAM?* Journal of the Association for Information Systems, 8, 211-218. Retrieved from http://aisel.aisnet.org/jais/vol8/iss4/16

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods, 1,* 80-91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Black, R. A., Yang, Y., Beitra, D., & McCaffrey, S. (2015). Comparing fit and reliability estimates of a psychological instrument using second-order CFA, bifactor, and essentially Tau-equivalent (coefficient alpha) models via Amos 22. *Journal of Psychoeducational Assessment, 33,* 451–472. https://doi.org/10.1177/0734282914553551

Boekaerts, M. (2017). Cognitive load and self-regulation: Attempts to build a bridge. *Learning and Instruction, 51,* 90–97. https://doi.org/10.1016/j.learninstruc.2017.07.001

Bong, M. (2001). Role of self-efficacy and task-value in predicting college students' course performance and future enrollment intentions. *Contemporary Educational Psychology, 26,* 553-570. https://doi.org/10.1006/ceps.2000.1048

Boucsein, W. (2012). *Electrodermal activity* (2nd Ed). New York: Springer.

Braithwaite, J., Watson, D., Robert, J., & Mickey, R. (2013). *A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments*. Retrieved from https://doi.org/10.1017.S0142716405050034

Bretz, F., Hothorn, T., & Westfall, P. (2016). *Multiple comparisons using R*. In Multiple Comparisons Using R. https://doi.org/10.1080/00401706.1964.10490181

Brouwer, A. M., Hogervorst, M. A., Holewijn, M., & van Erp, J. B. F. (2014). Evidence for effects of task difficulty but not learning on neurophysiological variables associated with effort. *International Journal of Psychophysiology, 93,* 242- 252. https://doi.org/10.1016/j.ijpsycho.2014.05.004

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated Cognition and the Culture of Learning. Educational Researcher, 18. 141-178. https://doi.org/10.3102/0013189X018001032

Burton-Jones, A., & Hubona, G. S. (2006). The mediation of external variables in the technology acceptance model. *Information and Management, 43,* 706-717. https://doi.org/10.1016/j.im.2006.03.007

Chang, C. C., & Yang, F. Y. (2010). Exploring the cognitive loads of high-school students as they learn concepts in web-based environments. *Computers and Education, 55,* 673-680. https://doi.org/10.1016/j.compedu.2010.03.001

Chen, K. C., & Jang, S. J. (2010). Motivation in online learning: Testing a model of self-determination theory. *Computers in Human Behavior, 28,* 31-50. https://doi.org/10.1016/j.chb.2010.01.011

Chen, O., & Kalyuga, S. (2020). *Cognitive Load Theory, Spacing Effect, and Working Memory Resources Depletion: Implications for Instructional Design.* In Form, Function, and Style in Instructional Design: Emerging Research and Opportunities. https://doi.org/10.4018/978-1-5225-9833-6.ch001

Chen, F., Zhou, J., Wang, Y., Yu, K., Arshad, S. Z., Khawaji, A., & Conway, D. (2016). *Robust Multimodal Cognitive Load Measurement.* In Human-Computer Interaction Series. Retrieved from https://link.springer.com/book/10.1007%2F978-3-319-31700-7

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior, 25,* 315-324. https://doi.org/10.1016/j.chb.2008.12.020

Clarebout, G., & Elen, J. (2006). Tool use in computer-based learning environments: towards a research framework. *Computers in Human Behavior, 22,* 389-411. https://doi.org/10.1016/j.chb.2004.09.007

Clarebout, G., & Elen, J. (2007). In Search of Pedagogical Agents' Modality and Dialogue Effects in Open Learning Environments. *E-Journal of Instructional Science and Technology.* Retrieved from https://files.eric.ed.gov/fulltext/EJ846724.pdf

Clarebout, G., Horz, H., Schnotz, W., & Elen, J. (2010). The relation between self-regulation and the embedding of support in learning environments. *Educational Technology Research & Development, 58,* 573-587. https://doi.org/10.1007/s11423-009-9147-4

Conger, D., & Dickson, L. (2017). Gender imbalance in higher education: Insights for college administrators and researchers. *Research in Higher Education, 58,* 214–230. https://doi.org/10.1007/s11162-016-9421-3

Conway, D., Dick, I., Li, Z., Wang, Y., & Chen, F. (2013). *The effect of stress on cognitive load measurement.* In Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-642-40498-6_58

Cook, D. A., & McDonald, F. S. (2008). E-learning: Is there anything special about the "e"? *Perspectives in Biology and Medicine, 51,* 5-21. https://doi.org/10.1353/pbm.2008.0007

Cranford, K N., Tiettmeyer, J.M., Chuprinko, B. C., Jordan, S., & Grove, N.P. (2014). Measuring load on working memory: the use of heart rate as a means of measuring chemistry students cognitive load. *Journal of Chemical Education, 91,* 641–647. https://doi.org/10.1021/ed400576n

Cukurova, M., Kent, C., & Luckin, R. (2019). Artificial intelligence and multimodal data in the service of human decision-making: A case study in debate tutoring. *British Journal of Educational Technology, 50,* 3032-3046 https://doi.org/10.1111/bjet.12829

Cuieford, J. P. (1965). *Fundamental statistics in psychology and education* (4th ed.). New York, NY: McGraw Hill.

Czerkwaski, C., & Leyman, E. (2016). An instructional design framework for fostering student engagement in online learning environments. *Technology Trends, 60,* 532-539. https://doi.org/10.1007/s11528-016-0110-z

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13,* 319-340. https://www.jstor.org/stable/249008

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science, 35.* https://doi.org/10.1287/mnsc.35.8.982

Davies, R., Nyland, R., Bodily, R., Chapman, J., Jones, B., & Young, J. (2017). Designing Technology-Enabled Instruction to Utilize Learning Analytics. *TechTrends, 61,* 155-161. https://doi.org/10.1007/s11528-016-0131-7

De Bruin, A. B. H., & van Merriënboer, J. J. G. (2017). Bridging cognitive load and self-regulated learning research: a complementary approach to contemporary issues in educational research. *Learning and Instruction, 51,* 1-9. https://doi.org/10.1016/j.learninstruc.2017.06.001

DeLeeuw, K. E., & Mayer, R. E. (2008). A Comparison of Three Measures of Cognitive Load: Evidence for Separable Measures of Intrinsic, Extraneous, and Germane Load. *Journal of Educational Psychology, 100,* 223-234. https://doi.org/10.1037/0022-0663.100.1.223

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning, 34*, 338–349. https://doi.org/10.1111/jcal.12288

Dindar, M., Malmberg, J. Järvelä, S., Haataja, E. & Kirschner (2019). Matching self-reports with electrodermal activity data: investigating temporal changes in self-regulated learning. *Education and Information Technologies. 1*-18. https://doi.org/10.1007/s10639-019-10059-5

Dollinger, S. J. (2000). Locus of control and incidental learning: an application to college student success. *College Student Journal, 34, 4,* 537–540. Retrieved from Retrieved from https://www.questia.com/library/journal/1G1-69750050/locus-of-control-and-incidental-learning-an-application

Duncan, T. G., & McKeachie, W. J. (2005). The making of the motivated strategies for learning questionnaire. *Educational Psychologist, 40,* 117-128. https://doi.org/10.1207/s15326985ep4002_6

Elen, J. (2004). Turning electronic learning environments into useful and influential "instructional design anchor points". *Educational Technology Research & Development, 52,* 67-73. https://doi.org/10.1007/BF02504719

Elen, J. (2015). *Technologie en onderwijs: onderwijskundige beschouwingen.* In Muziekpedagogiek in beweging. Technologie als medium, L. Nijs and T. De Baets (Eds.). België. Euprint Editions., Heverlee, 15-26.

Elen, J. (2020). “Instructional disobedience”: a largely neglected phenomenon deserving more systematic research attention. *Educational Technology Research and Development, 4,* 67-73. https://doi.org/10.1007/s11423-020-09776-3

European Commission. (2020). *Data protection in the EU.* Retrieved from https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en#:~:text=The%20European%20Commission%20has%20appointed%20a%20Data%20Protection,in%20cooperation%20with%20the%20European%20data%20protection%20supervisor

Evens, M., Elen, J., & Depaepe, F. (2017). Effects of opportunities to learn in teacher education on the development of teachers’ professional knowledge of French as a foreign language. *Journal of Advances in Education Research, 4,* 265-279 https://dx.doi.org/10.22606/jaer.2017.24007

Fishbein, M., & Ajzen, I. (1975). *Belief, attitudes, intention and behavior: an introduction to theory and research.* Reading, MA: Addison-Wesley.

Fichman, M., & Cummings, J. N. (2003). Multiple imputation for missing data: making the most of what you know. *Organizational Research Methods.* https://doi.org/10.1177/1094428103255532

Frerejean, J., van Merriënboer, J. J. G., Kirschner, P. A., Roex, A., Aertgeerts, B., & Marcellis, M. (2019a). Designing instruction for complex learning: 4C/ID in higher education. *European Journal of Education, 54,* 513-524. https://doi.org/10.1111/ejed.12363

Frerejean, J., Velthorst, G. J., van Strien, J. L. H., Kirschner, P. A., & Brand-Gruwel, S. (2019b). Embedded instruction to learn information problem solving: Effects of a whole task approach. *Computers in Human Behavior, 90,* 117-130. https://doi.org/10.1016/j.chb.2018.08.043

Frick, T. W., Chadha, R., Watson, C., Wang, Y., & Green, P. (2009). College student perceptions of teaching and learning quality. *Educational Technology Research and Development, 57.* 705–720. https://doi.org/10.1007/s11423-007-9079-9

Frick, T. W., Chadha, R., Watson, C., & Zlatkovska, E. (2010). Improving course evaluations to improve instruction and complex learning in higher education. *Educational Technology Research and Development, 58.* 115-136. https://doi.org/10.1007/s11423-009-9131-z

Gašević, D., Dawson, S., & Siemens, G. (2015). Let’s not forget: Learning analytics are about learning. *TechTrends, 59.* https://doi.org/10.1007/s11528-014-0822-x

Glogger-Frey, I., Gaus, K., & Renkl, A. (2017). Learning from direct instruction: Best prepared by several self-regulated or guided invention activities? *Learning and Instruction, 51, 25-35.* https://doi.org/10.1016/j.learninstruc.2016.11.002

Greene, J. A., & Azevedo, R. (2007). A theoretical review of Winne and Hadwin’s model of self-regulated learning: New perspectives and directions. *Review of Educational Research, 77,* 334-372. https://doi.org/10.3102/003465430303953

Greene, J. A., Moos, D. C., Azevedo, R., & Winters, F. I. (2008). Exploring differences between gifted and grade-level students' use of self-regulatory learning processes with hypermedia. *Computers & Education, 50*, 1069-1083. https://doi.org/10.1016/j.compedu.2006.10.004

Haapalainen, E., Kim, S., Forlizzi, J. F., & Dey, A. K. (2010). *Psycho-Physiological Measures for Assessing Cognitive Load.* In proceedings of the 12th ACM International Conference on Ubiquitous Computing.

Henrie, C. R., Bodily, R., Larsen, R., & Graham, C. R. (2018). Exploring the potential of LMS log data as a proxy measure of student engagement. *Journal of Computing in Higher Education, 30*, 344-362. http://dx.doi.org/10.1007/s12528-017-9161-1

Herbig, N., Düwel, T., Helali, M., Eckhart, L., Schuck, P., Choudhury, S., & Krüger, A. (2020). *Investigating Multi-Modal Measures for Cognitive Load Detection in E-Learning.* In Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '20). Association for Computing Machinery, New York, NY, USA, 88–97. https://doi.org/10.1145/3340631.3394861

Herborn, K.A., Graves, J.L., Jerem, P., Evans, N.P., Nager, R., McCafferty, D.J., & McKeegan, D.E.F. (2015). Skin temperature reveals the intensity of acute stress. *Physiology and Behavior, 1,* 225-230. https://doi.org/10.1016/j.physbeh.2015.09.032

Herrington, J., Oliver, R., & Reeves, T. C. (2003). Patterns of engagement in authentic online learning environments. *Australasian Journal of Educational Technology, 19,* 59-71. https://doi.org/10.14742/ajet.1701

Hoogerheide, V., Renkl, A., Fiorella, L., Paas, F., & van Gog, T. (2018). Enhancing Example-Based Learning: Teaching on Video Increases Arousal and Improves Problem-Solving Performance. *Journal of Educational Psychology, 211,* 45-56. https://doi.org/10.1037/edu0000272

Huang, H., Rauch, U., & Liaw, S. (2010). Investigating learners' attitudes toward virtual reality learning environments: based on a constructivist approach. *Computers & Education, 55,* 1171-1182. https://doi.org/10.1016/j.compedu.2010.05.014

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure modeling: conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55. http://dx.doi.org/10.1080/10705519909540118

Iani, C., Gopher, D., & Lavie, P. (2004). Effects of task difficulty and invested mental effort on peripheral vasoconstriction. *Psychophysiology, 41,* 789-798. https://doi.org/10.1111/j.1469-8986.2004.00200.x

Ikehara, C., & Crosby, M. (2005). *Assessing Cognitive Load with Physiological Sensors.* In Proceedings of the 38th Annual Hawaii International Conference on System Sciences, Big Island, HI, USA, 2005, pp. 295a-295a. DOI: 10.1109/HICSS.2005.103

Islam, N. (2013). Investigating e-learning system usage outcomes in the university context. *Computers & Education, 69,* 387-399. https://doi.org/10.1016/j.compedu.2013.07.037

Jaggers, S., & Xu, D. (2016). How do online course design features influence student performance? *Computers & Education, 95,* 270-284. https://doi.org/10.1016/j.compedu.2016.01.014

Jeong, H., & Hmelo-Silver, C. E. (2010). Productive use of learning resources in an online problem-based learning environment. *Computers in Human Behavior, 26,* 84-99. https://doi.org/10.1016/j.chb.2009.08.001

Jiang, L., Elen, J., & Clarebout, G. (2009). The relationships between learner variables, tool-usage behaviour and performance. *Computers in Human Behavior, 25,* 501-509. https://doi.org/10.1016/j.chb.2008.11.006

Johannessen, E., Szulewski, A., Radulovic, N., White, M., Braund, H., Howes, D., … Davies, C. (2020). Psychophysiologic measures of cognitive load in physician team leaders during trauma resuscitation. *Computers in Human Behavior, 111.* https://doi.org/10.1016/j.chb.2020.106393

Johnson, S. D., & Aragon, S. R. (2003). An instructional strategy framework for online learning environments. *New Directions for Adult and Continuing Education, 2003(100),* 31–43. https://doi.org/10.1002/ace.117

Jonassen, D. H. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology Research and Development, 45,* 65-94. https://doi.org/10.1007/bf02299613

Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development, 48,* 63-85. https://doi.org/10.1007/BF02300500

Juarez Collazo, N. A., Elen, J., & Clarebout, G. (2015). The multiple effects of combined tools in computer-based learning environments. *Computers in Human Behavior, 51,* 82-95. https://doi.org/10.1016/j.chb.2015.04.050

Juarez Collazo, N. A., Wu, X., Elen, J., & Clarebout, G. (2014). Tool use in computer-based learning environments: adopting and extending the Technology Acceptance Model. *Hindawi Publishing Corporation, 2014,* 1-10. https://doi.org/10.1155/2014/736931

Joo, Y. J., Lim, K. Y., & Kim, J. (2013). Locus of control, self-efficacy, and task value as predictors of learning outcome in an online university context. *Computers and Education. 62,* 149–158. https://doi.org/10.1016/j.compedu.2012.10.027

Kalyuga, S., & Singh, A. M. (2016). Rethinking the boundaries of cognitive load theory in complex learning. *Educational Psychology Review, 28,* 831-852. https://doi.org/10.1007/s10648-015-9352-0

Karthikeyan, P., Murugappan, M., & Yaacob, S. (2012). Descriptive analysis of skin temperature variability of sympathetic nervous system activity in stress. *Journal of Physical Therapy Science, 24*. 1341-1344. https://doi.org/10.1589/jpts.24.1341

Kim, H.G., Cheon, E.J., Bai, D.S., Lee, Y.H., & Koo, B.H. (2018). Stress and heart rate variability: A meta-analysis and review of the literature. *Psychiatry Investigation, 15,* 235-245. https://doi.org/10.30773/pi.2017.08.17

King, W. R., & He. J. (2006). A meta-analysis of the technology acceptance model. *Information & Management, 43,* 740-755. https://doi.org/10.1016/j.chb.2010.06.025

Kirschner, P. A., Ayres, P., & Chandler, P. (2011). Contemporary cognitive load theory research: The good, the bad and the ugly. *Computers in Human Behavior, 27,* 99–105. https://doi.org/10.1016/j.chb.2010.06.025

Kirschner, F., Kester, L., & Corbalan, G. (2011). Cognitive load theory and multimedia learning, task characteristics and learning engagement: The Current State of the Art. *Computers in Human Behavior, 27,* 1-4. https://doi.org/10.1016/j.chb.2010.05.003

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology, 8.* https://doi.org/10.3389/fpsyg.2017.01997

Kline, M. S. (2013). *Principles and practice of structural equation modeling.* (4th. ed.). The Guilford Press, New York.

Lane H.C., D'Mello S.K. (2019). *Uses of Physiological Monitoring in Intelligent Learning Environments: A Review of Research, Evidence, and Technologies*. In: Parsons T., Lin L., Cockerham D. (eds) Mind, Brain and Technology. Educational Communications and Technology: Issues and Innovations. Springer, Cham. https://doi.org/10.1007/978-3-030-02631-8_5

Lau, S., & Woods, P. C. (2009). Understanding learner acceptance of learning objects: the re-learning environments of learning object characteristics and individual differences. *British Journal of Educational Technology, 40,* 1059-1075. https://doi.org/10.1111/j.1467-8535.2008.00893.x

Larusson, J. A., & White, B. (2014). Learning analytics: From research to practice. *Technology, Knowledge, and Learning, 20.* https://doi.org/10.1007/978-1-4614-3305-7

LeCroy, C. W., & Krysik, J. (2007). Understanding and interpreting effect size measures. *Social Work Research, 31.* 243-248. https://doi.org/10.1093/swr/31.4.243

Lee, Y., Choi, J., & Kim, T. (2013). Discriminating factors between completers of and dropouts from online learning courses. *British Journal of Educational Technology, 44,* 328-337. https://doi.org/10.1111/j.1467-8535.2012.01306.x

Lee, J. Y., Donkers, J., Jarodzka, H., & van Merriënboer, J. J. G. (2019). How prior knowledge affects problem-solving performance in a medical simulation game: Using game-logs and eye-tracking. *Computers in Human Behavior, 99,* 268-277. https://doi.org/10.1016/j.chb.2019.05.035

Lee, B., Yoon, J., & Lee, I. (2009). Learners' acceptance of e-learning in South Korea: Theories and results. *Computers & Education, 53,* 1320–1329. https://doi.org/10.1016/j.compedu.2009.06.014

Legris, P., Ingham, J. & Collerette, P. (2003). Why do people use information technology? Critical review of the technology acceptance model. *Information & Management, 40,* 191-204. https://doi.org/10.1016/S0378-7206(01)00143-4

Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods, 45,* 1085-1072. https://doi.org/10.3758/s13428-013-0334-1

Liapis, A., Katsanos, C., Sotiropoulos, D., Xenos, M., & Karousos, N. (2015). *Recognizing emotions in human computer interaction: Studying stress using skin conductance.* In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 255-262. https://doi.org/10.1007/978-3-319-22701-6_18

Liaw, S. S., & Huang, H. M. (2013). Perceived satisfaction, perceived usefulness and interactive learning environments as predictors to self-regulation in e-learning environments. *Computers and Education, 60,* 14-24. https://doi.org/10.1016/j.compedu.2012.07.015

Liem, A. D., Lau, S., & Nie, Y. (2008). The role of self-efficacy, task value, and achievement goals in predicting learning strategies, task disengagement, peer relationship, and achievement outcome. *Contemporary Educational Psychology, 33,* 486-512. https://doi.org/10.1016/j.cedpsych.2007.08.001

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

Long, P., & Siemens, G. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review.* Retrieved from https://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education

Lust, G., Elen, J., & Clarebout, G. (2013). Students' tool-use within a web enhanced course: Explanatory mechanisms of students' tool-use pattern. *Computers in Human Behavior, 29,* 2013-2021. https://doi.org/10.1016/j.chb.2013.03.014

Lust, G., Juarez Collazo, N. A., Elen, J., & Clarebout, G. (2012). Content Management Systems: Enriched learning opportunities for all? Computers in *Human Behavior, 28,* 795-808. https://doi.org/10.1016/j.chb.2011.12.009

Margaryan, A., Bianco, M., & Littlejohn, A. (2015). Instructional quality of massive open online courses (MOOCs). *Computers & Education, 80,* 77–83. https://doi.org/10.1016/j.compedu.2014.08.005

Martens, R., Bastiaens, T. & Kirschner, P. A. (2007). New learning design in distance education: the impact on student perception and motivation. *Distance Education, 28,* 81-93. https://doi.org/10.1080/01587910701305327

Martens, R. L., Gulikers, J., & Bastiaens, T. (2004). The impact of intrinsic motivation on e-learning in authentic computer tasks. *Journal of Computer Assisted Learning. 20,* 368-376. https://doi.org/10.1111/j.1365-2729.2004.00096.x

Martin, S. (2014). Measuring cognitive load and cognition: metrics for technology-enhanced learning. *Educational Research and Evaluation, 20,* 592-621. https://doi.org/10.1080/13803611.2014.997140

Mayer, R. E. (2014). Incorporating motivation into multimedia learning. *Learning and Instruction, 29,* 171–173. https://doi.org/10.1016/j.learninstruc.2013.04.003

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*, 43-52. https://doi.org/10.1207/S15326985EP3801_6

McGill, T. J. & Klobas, J. E. (2009). A task-technology fit view of learning management system impact. Computers & Education, 52, 496-508. https://doi.org/10.1016/j.compedu.2008.10.002

Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development, 50,* 43-59. https://doi.org/10.1007/bf02505024

Milligan, S. K. (2018). *Methodological foundations for the measurement of learning in learning analytics.* In ACM International Conference Proceeding Series. https://doi.org/10.1145/3170358.3170391

Moissa B., Bonnin G., Boyer A. (2019). *Exploiting Wearable Technologies to Measure and Predict Students' Effort.* In: Buchem I., Klamma R., Wild F. (eds) Perspectives on Wearable Enhanced Learning (WELL). Springer, Cham. https://doi.org/10.1007/978-3-319-64301-4_19

Moos, D. C., & Azevedo, R. (2009). Self-efficacy and prior domain knowledge: To what extent does monitoring mediate their relationship with hypermedia learning? *Metacognition and Learning, 197-216.* https://doi.org/10.1007/s11409-009-9045-5

Morton, J., Vanneste, P., Larmuseau, C., Van Acker, B., Raes, A., Bombeke, K., Cornillie, F., Saldien, J., De Marez, L. (2019). *Identifying predictive EEG features for cognitive overload detection in assembly workers in Industry 4.0.* In Proceedings of the 3rd International Symposium on Human Mental Workload: Models and Applications (H-WORKLOAD 2019). Rome, Italy. https://arrow.tudublin.ie/hwork19/1/

Mudrick, N. V., Taub, M., Azevedo, R., Price, M. J., & Lester, J. (2017). *Can physiology indicate cognitive, affective, metacognitive, and motivational self-regulated learning processes during multimedia learning?* Paper presented at the Annual Meeting of the American Educational Research Association (AERA), San Antonio, TX.

National Academies of Sciences, Engineering, and Medicine. (2018). How People Learn II: Learners, Contexts, and Cultures. *The National Academies Press.* https://doi.org/10.17226/24783

Ng, W. (2015). *New digital technology in education: Conceptualizing professional learning for educators.* Retrieved from https://doi.org/10.1007/978-3-319-05822-1

Nguyen, Q., Huptych, M., & Rienties, B. (2018). *Linking students' timing of engagement to learning design and academic performance.* In ACM International Conference Proceeding Series. https://doi.org/10.1145/3170358.3170398

Nkambou, R. & Bourdeau, J. & Mizoguchi, R. (2010). Introduction: What Are Intelligent Tutoring Systems, and Why This Book? *Studies in Computational Intelligence. 308.* DOI: 10.1007/978-3-642-14363-2_1.

Noroozi, O., Pijeira-Díaz, H. J., Sobocinski, M., Dindar, M., Järvelä, S., & Kirschner, P. A. (2020). Multimodal data indicators for capturing cognitive, motivational, and emotional learning processes: A systematic literature review. *Education and Information Technologies.* https://doi.org/10.1007/s10639-020-10229-w

Nourbakhsh, N., Wang, Y., Chen, F., & Calvo, R. A. (2012). *Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks.* In Proceedings of the 24th Australian Computer-Human Interaction Conference, OzCHI 2012. https://doi.org/10.1145/2414536.2414602

Opfermann, M., Scheiter, K., Gerjets, P., & Schmeck, A. (2013). *Hypermedia and Self-Regulation: An Interplay in Both Directions.* In book: International handbook of metacognition and learning technologies. https://doi.org/10.1007/978-1-4419-5546-3_9

Or, C. K., & Duffy, V. G. (2007). Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement. Occupational Ergonomics, 7. 83–94. Retrieved from https://www.engineering.hku.hk/enggke/library/IMSE/CalvinOr/cor_8.pdf

Paas, F. (1992). Training strategies for attaining transfer of problem solving skills in statistics: A cognitive load approach. *Journal of Educational Psychology, 84,* 429–434. http://dx.doi.org/10.1037/0022-0663.84.4.429

Paas, F., Tuovinen, J. E., van Merriënboer, J. J. G., & Darabi, A. A. (2005). A Motivational Perspective on the Relation Between Mental Effort and Performance: Optimizing Learner Involvement in Instruction. *Educational Technology Research and Development, 53,* 25–34. https://doi.org/10.1007/BF02504795

Paas, F., Tuovinen, J., Tabbers, H., & Van Gerven, P. W. M. (2010). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist, 38,* 63-71. https://doi.org/10.1207/S15326985EP3801

Perkins, D. (1985). The fingertip effect: How information-processing technology shapes thinking. *Educational Researcher, 14,* 11–17. https://doi.org/10.3102/0013189X014007011

Phan, T., McNeil, S. G., & Robin, B. R. (2016). Students' patterns of engagement and course performance in a Massive Open Online Course. *Computers and Education, 95,* 36-44. https://doi.org/10.1016/j.compedu.2015.11.015

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and Self-Regulated Learning Components of Classroom Academic Performance. *Journal of Educational Psychology, 82,* 33–40. https://doi.org/10.1037/0022-0663.82.1.33

Raaijmakers, S. F., Baars, M., Schaap, L., Paas, F., & van Gog, T. (2017). Effects of performance feedback valence on perceptions of invested mental effort. *Learning and Instruction, 51,* 35-46. https://doi.org/10.1016/j.learninstruc.2016.12.002

Ramasubramanian, K., & Singh, A. (2016). *Machine learning using R.* https://doi.org/10.1007/978-4842-2334-5

Revere, L., & Kovach, J. V. (2011). Online technologies for engaged learning, a meaningful synthesis for educators. *The Quarterly Review of Distance education, 12,* 113-124. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.466.7571

Rienties, B. & Toetenel, L. (2016). The impact of learning design on student behaviour, satisfaction and performance: a cross-institutional comparison across 151 modules. *Computers in Human Behavior, 60,* 333-341. https://doi.org/10.1016/j.chb.2016.02.074

Rienties, B., Tempelaar, D., Van den Bossche, P., Gijselaers, W., & Segers, M. (2009). The role of academic motivation in Computer-Supported Collaborative Learning. *Computers in Human Behavior, 25,* 1195-1206. https://doi.org/10.1016/j.chb.2009.05.012

Rosseel, Y. (2012). Lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48,* 1-36. http://dx.doi.org/10.18637/jss.v048.i02

Roth, A. J. (1999). Multiple comparison procedures for discrete test statistics. *Journal of Statistical Planning and Inference, 82,* 101-117. https://doi.org/http://dx.doi.org/10.1016/S0378-3758(99)00034-8

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

Savalei, V., & Bentler, P. (2009). A two-stage approach to missing data: theory and application to auxiliary variables. *Structural Equation Modeling: A Multidisciplinary Journal, 16,* 477-497. doi: 10.1080/10705510903008238

Scharinger, C., Soutschek, A., Schubert, T., & Gerjets, P. (2015). When flanker meets the n-back: What EEG and pupil dilation data reveal about the interplay between the two central-executive working memory functions inhibition and updating. *Psychophysiology*. 1293-1304. https://doi.org/10.1111/psyp.12500

Sharma, K. & Giannakos, M. (2020). Multimodal data capabilities for learning: What can multimodal data tell us about learning? *British Journal of Educational Technology.* https://doi.org/10.1111/bjet.12993

Schepers, J., & Wetzels, J. (2007). A meta-analysis of the technology acceptance model: investigating subjective norm and moderating effects. *Information & Management, 44,* 90-103. https://doi.org/10.1016/j.im.2006.10.007

Selim, H. M. (2003). An empirical investigation of student acceptance of course websites. *Computers & Education, 40,* 343-360. https://doi.org/10.1016/S0360-1315(02)00142-2

Selim, H. M. (2007). Critical success factors for e-learning acceptance: Confirmatory factor models. *Computers & Education, 49,* 396-431. https://doi.org/10.1016/j.compedu.2005.09.004

Shi, Y., Ruiz, N., Taib, R., Choi, E., & Chen, F. (2007). *Galvanic skin response (GSR) as an index of cognitive load.* In CHI '07 extended abstracts on Human factors in computing systems - CHI '07. https://doi.org/10.1145/1240866.1241057

Schmeck, A., Opfermann, M., van Gog, T., Paas, F., & Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving: differences between immediate and delayed ratings. *Instructional Science, 43,* 93-114. https://doi.org/10.1007/s11251-014-9328-3

Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy, 2,* 419–436. https://doi.org/10.1207/S15327078IN0204_02

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: a review. *The Journal of Educational Research, 99,* 323-338. https://doi.org/10.3200/JOER.99.6.323-338

Schwaighofer, M., Bühner, M., & Fischer, F. (2017). Executive functions in the context of complex learning: Malleable moderators? *Frontline Learning Research, 5,* 58-75. https://doi.org/10.14786/flr.v5i1.268

Selwyn, N. What is the problem with learning analytics? *Journal of Learning Analytics, 6,* 11-19. http://dx.doi.org/10.18608/jla.2019.63.3

Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., & Ehlert, U. (2010). Discriminating stress from cognitive load using a wearable eda device. *IEEE Transactions on Information Technology in Biomedicine, 14.* 410-417. https://doi.org/10.1109/TITB.2009.2036164

Smets, E., Rios Velazquez, E., Schiavone, G., Chakroun, I., D'Hondt, E., De Raedt, W., … Van Hoof, C. (2018a). Large-scale wearable data reveal digital phenotypes for daily-life stress detection. *Npj Digital Medicine*. https://doi.org/10.1038/s41746-018-0074-9

Smets, E., Schiavone, G., Velazquez, E. R., De Raedt, W., Bogaerts, K., Van Diest, I., & Van Hoof, C. (2018b). Comparing task-induced psychophysiological responses between persons with stress-related complaints and healthy controls: A methodological pilot study. *Health Science Reports.* https://doi.org/10.1002/hsr2.60

Slavin, R. (2003). *Educational psychology: Theory and practice.* Boston: Pearson Education.

Song, L., Singleton, E. S., Hill, J. R., & Koh, H. M. (2004). Improving online learning: student perceptions of useful and challenging characteristics. *Internet and Higher Education, 7,* 59-70. https://doi.org/10.1016/j.iheduc.2003.11.003

Spikol, D. & Cukurova, M. (2019). *Multimodal Learning Analytics.* In book: Encyclopedia of Education and Information Technologies. Publisher: Springer, Cham.

Šumak, B., Hericko, M. &, Pušnik, M. (2011). A meta-analysis of e-learning technology acceptance: The role of user types and e-learning technology types. *Computers in Human Behavior, 27,* 2067-2077. https://doi.org/10.1016/j.chb.2011.08.005

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4,* 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22,* 123-138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J., van Merrienboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 3*, 251-196. https://doi.org/10.1023/A:1022193728205

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review, 32*, 261-292. https://doi.org/10.1007/s10648-019-09465-5

Siemens, G., & Baker, R. S. J. D. (2012). *Learning analytics and educational data mining: Towards communication and collaboration.* In ACM International Conference Proceeding Series. https://doi.org/10.1145/2330601.2330661

Song, H. S., Kalet, A. L., & Plass, J. L. (2016). Interplay of prior knowledge, self-regulation and motivation in complex multimedia learning environments. *Journal of Computer Assisted Learning, 32,* 31-50. https://doi.org/10.1111/jcal.12117

Stone, J.M., & Towse, J.N. (2015). A working memory test battery: Java-based collection of seven working memory tasks. *Journal of Open Research Software, 3.* https://doi.org/10.5334/jors.br

Taelman, J., Vandeput, S., Vlemincx, E., Spaepen, A., & Van Huffel, S. (2011). Instantaneous changes in heart rate regulation due to mental load in simulated office work. *European Journal of Applied Physiology, 111.* 1497-1505. https://doi.org/10.1007/s00421-010-1776-0

Taub, M., Azevedo, R., Bouchet, F., & Khosravifar, B. (2014). Can the use of cognitive and metacognitive self-regulated learning strategies be predicted by learners' levels of prior knowledge in hypermedia-learning environments? *Computers in Human Behavior, 39,* 356-367. https://doi.org/10.1016/j.chb.2014.07.018

Taipjutorus, W., Hansen, S., & Brown, M. (2012). Investigating a Relationship between Learner Control and Self-Efficacy in an Online Learning Environment. *Journal of Open, Flexible and Distance Learning, 16,* 56-69. Retrieved from https://files.eric.ed.gov/fulltext/EJ1079899.pdf

Tarhini, A., Hone, K., & Xiaohui, L. (2013). Factors affecting students' acceptance of e-learning environment in developing countries: a structural equation modeling approach. *International Journal of Information and Educational Technology, 3,* 54-59. DOI: 10.7763/IJIET.2013.V3.233

Teo, T. (2009). "Is there an attitude problem? Reconsidering the re-learning environment of attitude in TAM". *British Journal of Educational Technology, 40,* 1139-1141. https://doi.org/10.1111/j.1467-8535.2008.00913.x

Terras, M. M., & Ramsay, J. (2015). Massive open online courses (MOOCs): insights and challenges from a psychological perspective. *British Journal of Educational Technology, 46,* 472-487. https://doi.org/10.1111/bjet.12274

Thompson, R., Higgins, C. A., & Howell, J. M. (1991). Personal computing: toward a conceptual model of utilization. *MIS Quarterly, 15,* 125-143. DOI:10.2307/249443

Tsai, C.W. (2013). An effective online teaching method: The combination of collaborative learning with initiation and self-regulation learning with feedback. *Behaviour and Information Technology, 32,* 712–723. https://doi.org/10.1080/0144929X.2012.667441

Turner, M., Kitchenham, B., Brereton, P., Charters, S., & Budgen, D. (2010). Does the technology acceptance model predict actual use? A systematic literature review. *Information and Software Technology, 52,* 463-479. https://doi.org/10.1016/j.infsof.2009.11.005

Väljataga, T., & Laanpere, M. (2010). Learner control and personal learning environment: A challenge for instructional design. *Interactive Learning Environments, 18,* 277-291. https://doi.org/10.1080/10494820.2010.500546

van Gog, T., Sluijsmans, D. M. A., Brinke, D. J. ten, & Prins, F. J. (2010). Formative assessment in an online learning environment to support flexible on-the-job learning in complex professional domains. *Educational Technology Research and Development, 58,* 311-324. https://doi.org/10.1007/s11423-008-9099-0

Van Laer, S., & Elen, J. (2017). In search of attributes that support self-regulation in blended learning environments. *Education and Information Technologies, 22*, 1395–1454 https://doi.org/10.1007/s10639-016-9505-x

Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: a four-component instructional design model for technical training.* Englewood Cliffs, NJ: Educational Technology Publications.

van Merriënboer, J. J. G. (2013). Perspectives on problem solving and instruction. *Computers and Education, 64,* 153-160. https://doi.org/10.1016/j.compedu.2012.11.025

Van Merrienboer, J. J. G., Clark, R. E., & De Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational Technology Research and Development, 50,* 39-61. https://doi.org/10.1007/bf02504993

Van Merriënboer, J. J. G. van, & Kirschner, P. A. (2018). Ten Steps to Complex Learning: a Systematic Approach to Instruction and Instructional Design. *TechTrends, 62,* 204–205. https://doi.org/10.1007/s11528-018-0254-0

van Merrienboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the Load Off a Learner's Mind: Instructional Design for Complex Learning. *Educational Psychologist, 38,* 5–13. https://doi.org/10.1207/S15326985EP3801_2

Van Merriënboer, J. J. G., & Sluijsmans, D. M. A. (2009). Toward a synthesis of cognitive load theory, four-component instructional design, and self-directed learning. *Educational Psychology Review, 21,* 55-66. https://doi.org/10.1007/s10648-008-9092-5

Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17,* 147-177. https://doi.org/10.1007/s10648-005-3951-0

Vanneste, P., Raes, A., Morton, J., Bombeke, K., Van Acker, B., Larmuseau, C., Depaepe, F. & Van Den Noortgate. W. (2020). Towards measuring cognitive load during assembly work through multimodal physiological data. *Cognition, Technology & Work.* https://doi.org/10.1007/s10111-020-00641-0

Van Raaij, E., M. & Schepers, J. J. L. (2008). The acceptance and use of a virtual learning environment in China. *Computers & Education, 50,* 838-852. https://doi.org/10.1016/j.compedu.2006.09.001

Van Seters, J. R., Ossevoort, M. A., Tramper, J., & Goedhart, M. J. (2012). The influence of student characteristics on the use of adaptive e-learning material. *Computers and Education, 3,* 942-952. https://doi.org/10.1016/j.compedu.2011.11.002

Venkatesh, V., & Davis, F. D. (2000). Theoretical extension of the Technology Acceptance Model: Four longitudinal field studies. *Management Science, 46,* 186-204. https://doi.org/10.1287/mnsc.46.2.186.11926

Vinkers, C. H., Penning, R., Hellhammer, J., Verster, J. C., Klaessens, J. H. G. M., Olivier, B., & Kalkman, C. J. (2013). The effect of stress on core and peripheral body temperature in humans. *Stress, 16,* 520-520. https://doi.org/10.3109/10253890.2013.807243

Whipp, J. L., & Chiarelli, S. (2004). Self-regulation in a web-based course: A case study. *Educational Technology Research and Development, 52,* 5-22. https://doi.org/10.1007/bf02504714

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25,* 68-81. https://doi.org/10.1006/ceps.1999.1015

Wise, A., & Schaffer, D. (2015). Why theory matters more than ever in the age of big data. *Journal of Learning Analytics, 2,* 5-13. https://doi.org/10.18608/jla.2015.22.2

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models. *Educational and Psychological Measurement, 73,* 913–934. https://doi.org/10.1177/0013164413495237

Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., Baars, de Koning & Paas, F. (2019). *Educational Theories and Learning Analytics: From Data to Knowledge*. In Utilizing Learning Analytics to Support Study Success. https://doi.org/10.1007/978-3-319-64792-0_1

Worsley, M. (2018). *(Dis)Engagement matters: Identifying efficacious learning practices with multimodal learning analytics.* In ACM International Conference Proceedings of the 8th International Conference on Learning Analytics and Knowledge. 365–369. https://doi.org/10.1145/3170358.3170420

Yang, M., Shao, Z., Liu, Q. & Liu, C. (2017). Understanding the quality factors that influence the continuance intention of students toward participation in MOOCs. *Educational Technology Research and Development, 65,* 1195–1214. https://doi.org/10.1007/s11423-017-9513-6

Yelizarov, A., & Gamayunov, D. (2014). *Adaptive visualization interface that manages user's cognitive load based on interaction characteristics.* In ACM International Conference Proceeding Series. https://doi.org/10.1145/2636240.2636844

Younis Alsabawy, A., Cater-Steel, A. & Soar, J. (2016). Determinants of perceived usefulness in online learning systems. *Computer in Human Behavior, 64,* 843-858. https://doi.org/10.1016/j.chb.2016.07.065

Yousoof, M., & Sapiyan, M. (2013). Measuring cognitive load for visualizations in learning computer programming-physiological measures. *Ubiquitous and communication journal, 8.* 1410-1426. https://pdfs.semanticscholar.org/bdbd/8af1870e956e4a727e2449897077266fa8e5.pdf

Zagermann, J., Pfeil, U., & Reiterer, H. (2016). *Measuring Cognitive Load Using Eye Tracking Technology in Visual Computing.* In Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization. https://doi.org/10.1145/2993901.2993908

Zheng, R., & Cook, A. (2012). Solving complex problems: A convergent approach to cognitive load measurement. *British Journal of Educational Technology, 43,* 233-246. https://doi.org/10.1111/j.1467-8535.2010.01169.x

Zimmerman, B. J. (2000). Self-Efficacy: An Essential Motive to Learn. Contemporary Educational Psychology, 25. 82-91. https://doi.org/10.1006/ceps.1999.1016

Appendix: Abbreviations

AI	Artificial Intelligence
CLT	Cognitive Load Theory
CK	Content Knowledge
ECG	ElectroCardioGraphy
EDA	Electrodermal Activity
EEG	ElectroEncephaloGraphy
EDM	Educational Data Mining
GSR	Galvanic Skin Response
HR	Heart Rate
HRV	Heart Rate Variability
IT	Information Technology
ITS	Intelligent Tutoring Systems
LA	Learning Analytics
LAK	Learning Analytics and Knowledge community
LMS	Learning Management System
MMLA	Multimodal Learning Analytics
MOOC	Massive Open Online Courses
OSPAN	Computer Based Operation Span Test
PCK	Pedagogical Content Knowledge
PEOU	Perceived Ease of Use
PU	Perceived Usefulness
SCL	Skin Conductance Level
SNS	Sympathetic Nervous System
ST	Skin Temperature
VLE	Virtual Learning Environment
4C/ID	Four Component Instructional Design Model