Adapting Exercise Selection to Performance, Effort and Self-Esteem

Juliet A. Okpo[a], Judith Masthoff[a], Matt Dennis[b] and Nigel Beacham[a]

[a]University of Aberdeen, UK; [b]University of Portsmouth, UK

ABSTRACT

Adapting to learner characteristics is essential when selecting exercises for learners in an intelligent tutoring system. This paper investigates how humans adapt next exercise selection (in particular difficulty level) to learner personality, invested mental effort, and performance to inspire an adaptive exercise selection algorithm. First, the paper describes the investigations to produce validated materials for the main studies, namely the creation and validation of self-esteem personality stories, mental effort statements, and mathematical exercises with varying levels of difficulty. Next, through empirical studies, we investigate the impact on exercise selection of learner's self-esteem (low versus high self-esteem) and effort (minimal, little, moderate, much, and all possible effort). Three studies investigate this for learners who had different performances on a previous exercise: just passing, just failing, and performed well. Participants considered a fictional learner with a certain performance, self-esteem and effort, and selected the difficulty level of the next mathematical exercise. We found that self-esteem, mental effort, and performance all impacted the difficulty level of the exercises selected for learners. Finally, using the results from the studies, we propose an algorithm that selects exercises with varying difficulty levels adapted to learner characteristics.

KEYWORDS
Exercise Selection, Adaptation, Personality, Self-esteem, Mental effort,
Performance, Cognitive Efficiency

1. Introduction

This paper investigates how exercise selection (in particular difficulty-level) should be adapted to a learner's personality (self-esteem), performance, and effort. Selecting the next exercise for learners to do during any tutoring process is key in order to enhance the learner's competence in that domain and also to improve learner confidence in the domain (Camp, Paas, Rikers, & van Merrienboer, 2001; Corbalan, Kester, & van Merriënboer, 2008; Corbalan, Kester, & Van Merrienboer, 2011; Kostons, van Gog, & Paas, 2010; Salden, Paas, & Van Merriënboer, 2006; Stephens et al., 2013). Tutors from varied domains have used several combinations of learner and exercise characteristics (Okpo, 2016) in an attempt to facilitate the process of selecting the next tutoring exercise for learners. It is expected that to achieve success in the effective selection of exercises for learner's interest and cognitive abilities, an Intelligent Tutoring System should provide the mapping between the knowledge component of the learner and

CONTACT J. A. Okpo. Email: r02jao15@abdn.ac.uk

the exercise to be learned (Ravi & Sosnovsky, 2013). Another expectation is that learners should be involved in learning activities that are tailored and adapted to their knowledge level (Grivokostopoulou, Perikos, & Hatzilygeroudis, 2017). Our focus in this paper is on the selection of exercises based on learner personality and cognitive efficiency. Cognitive efficiency (CE) by most conceptions includes a performance and mental effort component (Camp et al., 2001; Van Gog, Kester, & Paas, 2011). Mental effort is perceived as the use of human inner resources to accomplish a given task (Paas & Van Merriënboer, 1993; Xie & Salvendy, 2000) and performance in our context depicts how well a learner has done in carrying out an academic task.

Personality has been the major learner characteristic in our various studies (Okpo, 2016; Okpo, Dennis, Masthoff, Smith, & Beacham, 2016; Okpo, Dennis, Smith, Masthoff, & Beacham, 2016; Okpo, Masthoff, Dennis, & Beacham, 2017; Okpo, Masthoff, Dennis, Beacham, & Ciocarlan, 2017). Our need to establish the personality traits that are important for exercise selection has led us to investigate how to convey a learner's personality to our participants who will select exercises. One option was to make participants have some form of interaction with the learner in order for them to know the learner's personality. However, this would not only need much time to be spent with the learner, it would also be hard to ensure that learners were only perceived to differ on the personality trait under investigation, and not on other characteristics such as affective state. Another option was to inform participants of the learner's personality explicitly, but such as short mention may not produce the empathy with the learner we required.

Based on these considerations, we decided to develop personality-trait stories for self-esteem following a similar approach to Dennis, Masthoff, and Mellish (2012) (see Section 3).

Another consideration was the type of exercises we needed for our investigations in exercise selection. We ran two pilot studies using exercises from the UK standard measure for giving exercises to learners in schools. From the results obtained (Okpo, Dennis, Smith, et al., 2016), it was discovered that the difficulty of each exercise compared with the other was not explicit enough and that perhaps our exercise difficulty levels were a bit too coarse grained (i.e. participants may have wanted to make more subtle changes in difficulty level than our difficulty levels allowed) We therefore needed a set of exercises with gradual and consistent changes in difficulty. This led us to the design of a study that validated the difficulty levels of a set of mathematical exercises (see Section 3). We choose mathematical exercises as mathematics possesses a well documented progression of procedure.

This paper is organized as follows. Section 2 provides the background, related work and describes a conceptual framework for adaptive exercise selection. Section 3 describes the development of the study materials which includes the construction and validation of a set of exercises with varying difficulties, the validation of stories that conveyed learner's self-esteem and the construction of mental effort statements. Section 4 describes the study design for investigating the impact of self-esteem, and cognitive efficiency on the selection of exercises for learners. Here we describe the methodology that was used for our studies. The results for the studies are presented in Sections 5 to 7. Section 8 combines the data of the three studies in order to analyse the impact of performance and possible interaction effects between performance, effort and self-esteem. Based on these results, in section 9 a regression analysis is performed resulting in an algorithm for next exercise selection. Finally, section 10 concludes the paper, discusses its limitations and provides suggestions for future work.
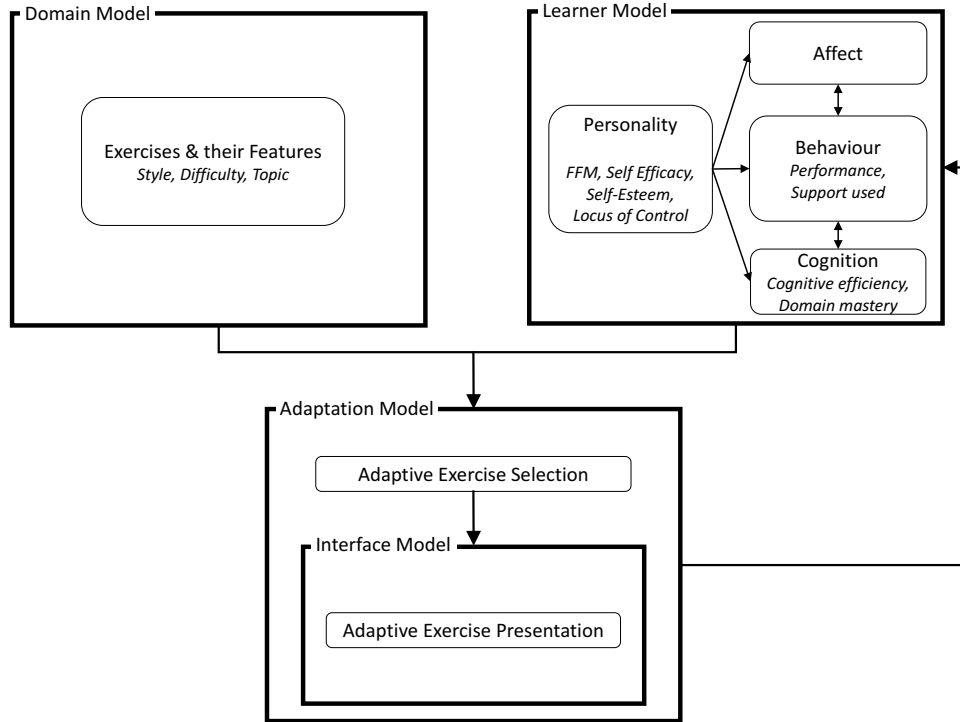
Figure 1. Conceptual Framework for Adaptive Exercise Selection

2. Background and Related Work

There has been much research on adapting learning content to different learner characteristics. Table 1 categorizes these learning characteristics and provides examples of existing ITS research which investigates adaptations to these categories. Furthermore, in six focus group studies (Okpo, Dennis, Masthoff, et al., 2016), we investigated the learner and exercise characteristics that could be considered when selecting the next exercises for learners by Intelligent Tutoring Systems. Participants felt that learner performance, personality, learning styles, age, knowledge and experience were most important, as well as exercise difficulty and the feedback exercises provide.

We believe that for better understanding of adaptive exercise selection for an Intelligent Tutoring System, a framework of the relationship of all the components for the system should be adequately represented and understood. We therefore attempt to define these concepts as they relate to exercise selection in intelligent tutoring.

Inspired by the literature, Figure 1 shows our conceptual framework for adaptive exercise selection. In this section, we describe this framework to provide context for the studies presented later which will focus on adaptation of exercise difficulty to self-esteem and cognitive efficiency.

The framework builds upon existing research on adaptive tutoring systems (A. T. Corbett, Koedinger, & Anderson, 1997) which uses four major components: the *domain model*, *learner model*, *interface model* and *adaptation model*.

Table 1. Categories of Learner Characteristics in Adaptive Learning Environments

Category	Sub category	Characteristic	Adaptive learning examples
Cognition	Style	Cognitive style	Assis et al. (2008); Lo, Chan, and Yeh (2012); Mampadi, Chen, Ghinea, and Chen (2011); Tennyson (1993); Tennyson, Thurlow, and Breuer (1987); Triantafillou, Pomportsis, Demetriadis, and Georgiadou (2004)
		Learning style	Conlan, Dagger, and Wade (2002); Latham, Crockett, McLean, and Edmonds (2012a, 2012b); Magoulas, Papanikolaou, and Grigoriadou (2003); Martinez and Bunderson (2000); Sun and Cheng (2007)
		Learning pattern	Jugo, Kovacic, and Slavuj (2014)
	Knowledge and skills	Episodic knowledge	Brusilovsky, Schwarz, and Weber (1996)
		Problem solving skill	Pholo and Ngwira (2013)
		Knowledge state/Domain mastery	Koedinger and Anderson (1993); Melis and Andres (2005); Mitrović, Djordjević-Kajan, and Stomenov (1996); Petrovica (2013); Ray and Belden (2007); Salden, Paas, and Van Merriënboer (2006); Shute (1995)
		Logical ability	Brusilovsky et al. (1996)
		Prior knowledge	Pon-Barry, Schultz, Bratt, Clark, and Peters (2006); Schwonke, Hauser, Nückles, and Renkl (2006)
		Knowledge assessment	Conati (2009)
	(other)	Mental Effort	Corbalan et al. (2008); Salden, Paas, and van Merriënboer (2006); Salden, Paas, and Van Merriënboer (2006)
		Degree of concentration	Koutsojannis, Beligiannis, Hatzilygeroudis, Papavlasopoulos, and Prentzas (2007); Tseng, Chu, Hwang, and Tsai (2008)
		Working memory capacity	Graf, Lin, et al. (2008); Lusk et al. (2009)
Affect		Affective states	Forbes-Riley, Rotaru, and Litman (2008); Graesser, Jeon, and Dufty (2008); Reategui, Boff, and Campbell (2008); Tennyson et al. (1987)
		Learner motivation	Beal and Lee (2005); del Solato and Du Boulay (1995); Dennis, Masthoff, and Mellish (2016); Fazlollahtabar and Mahdavi (2009); Montazemi and Wang (1995)
Behaviour	Support used	Hints obtained	Arevalillo-Herráez et al. (2014)
		Instructional support used	Aleven, Mclaren, Roll, and Koedinger (2006); Corbalan et al. (2008); Koutsojannis et al. (2007); Miwa et al. (2014)
	Performance	Learner progress	Hospers (2003); Verdú et al. (2012)
		Education background	Kelly (2008)
		Learning competence	Cheng, Shen, and Basu (2008); Corbalan et al. (2008); Davidovic, Warren, and Trichina (2003); Pon-Barry et al. (2006); Tseng et al. (2008)
		Number of tries	Hospers (2003)
		Learner errors	Mitrovic, Martin, and Mayo (2002); Oppermann and Rasher (1997),
		Learner responses	Dreher, Reiners, Dreher, and Dreher (2009); Kelly (2008); Mitrović et al. (1996); Mitrovic et al. (2002)
Personality		Self-Efficacy	Mcquiggan, Mott, and Lester (2008)
		Big 5	Dennis et al. (2016); Tsiriga and Virvou (2004)
		Self-Esteem	Okpo, Dennis, Smith, et al. (2016)
Other		Learner demographics/culture	Dreher et al. (2009); Read, Kahler, Strong, and Colder (2006); Reategui et al. (2008)

2.1. Domain model

In general, in an ITS, the *Domain Model* describes taught instructional content as well as the relationship between the domain contents. For exercise selection, the domain model contains the exercises and their features such as difficulty level, topic and style. The studies in this paper will focus on difficulty level, keeping topic and style constant.

2.1.1. Exercise difficulty

Exercises given to learners should be of the correct difficulty level, thereby ensuring that learners are focused in their engagement in the learning activities (Mora, Gallego-Durán, Molina-Carmona, & Llorens-Largo, 2017) without any obstruction in learning progression. Difficulty is a concept that depicts the application of a reasonable amount of effort for something to be accomplished (Nicholls & Miller, 1983; Varela, Thompson, & Rosch, 2017). Estimating difficulty is a complex process, hence the small number of studies that consider difficulty in learning adaptation processes.

Previous research has investigated the estimation of exercise difficulty level using different approaches. Quiz difficulty levels have been evaluated using similarity measures (Lin, Liu, Pang, & Wang, 2015). A graph based strategy for difficulty level estimation for chemistry was used by Wu and Cheng (2007). Foteini et al presented a neuro-fuzzy approach for the difficulty estimation of exercises on search algorithms (Grivokostopoulou, Perikos, & Hatzilygeroudis, 2015; Grivokostopoulou et al., 2017). In this method, specific characteristics of the exercises are taken as inputs to provide the difficulty of the exercises as output. In another study, the difficulty level of an exam was based on an item analysis approach where an item's difficulty was based on the proportion of people who correctly answered a test item, with the higher the proportion of those that answered the test item correctly, the lower the difficulty (Escudero, Reyna, & Morales, 2000).

In this paper, we will not investigate how to automatically detect exercise difficulty, but validate exercise difficulty through human studies to allow us to use these validated difficulties in our main studies. So, for our investigations, we designed a study to estimate the difficulty level of a set of mathematical exercises. Details of this study are reported in section 3 of this paper.

2.1.2. Exercise learning topic

Exercises will be available for different learning topics. For example, an exercise about mathematics differs from an exercise about learning the vocabulary of a foreign language. Even within a relatively well-defined topic, there can be differences in learning outcomes associated with exercises. For example, when learning say 10 foreign words, one could use so-called recognition or recall exercises. These do not only differ on difficulty but also on the learning outcomes achieved (for example, being able to recognize the correct translation of a foreign word, or being able to produce that translation oneself).

2.1.3. Exercise style

Different exercises styles may exist for the same learning topic. For example, exercises can be more verbal or visual, active or reflective, social or individual, sequential or global, sensory or intuitive (Alhathli, Masthoff, & Siddharthan, 2017). An adaptive exercise selection algorithm could potentially take this into account, for example

selecting the style best suited to a learner's personality, cognitive or learning style (Alhathli, Masthoff, & Siddharthan, 2016; Alhathli et al., 2017).

2.2. Learner Model

The *Learner model* contains the learner characteristics as well as the general behaviour of the learner within the system. It not only monitors the learner's behaviour in the system, but also updates other learner characteristics such as affect and mental effort. Our learner model contains four elements: Personality, Cognition, Affect and Behaviour.

2.2.1. Personality

Personality is a psychological construct which describes human behaviours in terms of measurable individual characteristics (Vinciarelli & Mohammadi, 2014). There are many different theories used by psychologists to measure and describe the personality of an individual – in this summary we describe constructs from the *trait* and *social learning* theories as these are those most commonly used by ITS researchers. These are the Five Factor Model, the leading Trait theory, and Self-Esteem, Self-Efficacy and Locus of Control from Social Learning Theory.

Five Factor Model: The five factor model (FFM) of personality (also known as the 'Big Five') (Goldberg, 1980) is a scientifically robust and complete measure of an individual's personality (McCrae & John, 1992). The dimensionality of the FFM cuts across all cultures (McCrae & John, 1992) and remains stable over time (Costa & McCrae, 1992; Judge, Higgins, Thoresen, & Barrick, 1999; Soldz & Vaillant, 1999). Personality is expressed as scores across the five traits: Extraversion (how talkative, energetic, assertive); Agreeableness (how good natured, cooperative, trustful); Conscientiousness (how orderly, responsible, dependable); Emotional Stability (how calm vs neurotic); Openness to Experience (how intellectual, imaginative, independent minded).

Self-Esteem: Based in social learning theory, self-esteem is defined as how favourably a person regards themselves (Rosenberg, 1986). For a learner to achieve better learning outcomes in a specific domain, they must believe in their abilities and this belief in the fact that they can produce a favourable outcome will in turn serve as motivation to learn. Self-esteem is seen as an important component of personality (Maslow, 1973), and is one of the most widely studied personality concepts in psychology (Judge, Erez, Bono, & Thoresen, 2002)

Significant associations have been found between self-esteem and all personality traits such as openness, conscientiousness, extraversion, agreeableness and neuroticism (Goldberg, 1980).

Self-efficacy: Self efficacy describes confidence in one's abilities (Bandura, 1986). In line with Bandura's social cognitive theory (Bandura, 1986), students' confidence in the performance of academic tasks can predict their ability to be successful (Pajares, 1996). These beliefs have also been hypothesized to influence other determinants of learning outcomes such as competence, past achievements, and skills. Confident learners usually exhibit a sense of responsibility for their learning thereby reducing boredom and distraction during learning. A learners confidence in their own ability to accomplish certain tasks is termed *self-efficacy*, and adaptive exercise selection may aim to boost learner's confidence, for learners with low self-efficacy.

Locus of control: An individual's Locus of Control represents the extent to which a person believes they can control events that affect them (Rotter, 1966). A learner with an internal locus of control believes that they control their own fate, i.e. they feel responsible for their academic performance and achievements. A learner with external locus of control believes that their fate is determined by external forces i.e. they believe that their academic performance is a result of how difficult a task was, the type of task given or how well they have been taught.

The studies in this paper focus on self-esteem as a starting point, however, other personality traits may be investigated in future work.

2.2.2. Cognition

As can be seen in Table 1, there are many aspects to cognition relevant for adapting ITS. Here, we will focus on cognitive efficiency and domain mastery.

Cognitive Efficiency is the amount of mental effort invested in a certain task combined with the quality of the observed performance (Taminiau et al., 2013). It is calculated using learner performance and learner mental effort (Camp et al., 2001). Mental effort is the use of human inner resources to accomplish a given task (Xie & Salvendy, 2000), so is an indicator of the load imposed on the learner's mental capacity by a task (Sweller, Ayres, & Kalyuga, 2011). The mental effort invested in doing exercises can be seen as the total amount of cognitive processing, which is regarded the cognitive cost of learning (Paas & Van Merriënboer, 1993). Mental effort invested in doing a task aligns closely with the perceived task complexity (Hoffman et al., 2016). There is also a clear relationship with work in educational psychology in which the efficiency matrix of instructional materials was formulated as a combination of the measures of mental load and task performance (Paas & Van Merriënboer, 1993). Performance (which is part of the behavioural characteristics, see below) has proved to be good for determining learning outcomes. For the studies in this paper, we will used performance and mental effort statements to represent cognitive efficiency. A study on the validation of mental effort statements is reported in section 3.3.

Domain mastery: According to Bloom (1956), there is a gradual progression of mastery. Therefore, teaching content such as exercises tend to be presented in stages and in a gradual progression from easy to difficult. As learners work on gradually more complex tasks, it enhances their understanding of the solution strategies. However, how difficult a task is, depends on factors affecting both the learner and the task. Learners with more mastery of a task will invest less mental effort in performing the task. Learning tasks selected must be at the right cognitive level for the learner, meaning that the tasks administered to the learners must neither be too easy, as this could bore the learner due to the lack of challenge in the learning content, nor too difficult, as this could overwhelm the learner due to excessive cognitive load (cf. the concept of flow (Csikszentmihalyi, 1988)). The right cognitive level depends on a model of the learner's domain mastery complemented by an observation of a learner's recent performance.

There is a relationship between personality and cognition. For example, it has been shown that personality could predict mental well-being (Malkoç, 2011) which in turn is positively associated with cognitive abilities (Llewellyn, Lang, Langa, & Huppert, 2008). A meta-analysis of studies has shown that personality is related to cognitive abilities Poropat (2009). For example, Rammstedt, Danner, and Martin (2016) found a

positive association between emotional stability and cognitive ability (for both verbal and numerical ability), whilst finding a negative association between conscientiousness and cognitive ability. Moutafi, Furnham, and Paltiel (2005) found that Openness positively predicted numerical reasoning, whilst Extraversion negatively predicted verbal, numerical, and abstract reasoning. Their results for conscientiousness and emotional stability were similar to those of Rammstedt et al. (2016). There is also a relationship between cognition and performance (Corno & Rohrkemper, 1985; Pintrich & De Groot, 1990).

2.2.3. Affect

A learner's affective state is a measure of the emotions they are experiencing at the current time. Affective states are often characterized by two dimensions: arousal (from calming to exciting) and valence (from highly positive to highly negative) (Russell, 1980). The popular PANAS scale defines positive affective states as: active, alert, attentive, determined, enthusiastic, proud and strong; and negative affect: afraid, scared, nervous, jittery, irritable, hostile, guilty, ashamed, upset and distressed (Watson, A, & Tellegen, 1998). Affective states such as attentiveness have been shown to be related to personality and conscientious individuals are likely to exhibit guilt when they fail to meet goals (Watson, Wiese, Vaidya, & Tellegen, 1999). It is likely that learner affect may be important for adaptive exercise selection in a future ITS, hence its inclusion in Figure 1. However, as this is not the current focus of our research, we do not elaborate further in this paper.

2.2.4. Behaviour

Performance: For adaptive exercise selection, performance describes how well a learner did on previous exercises (or tests), e.g. mistakes made and time taken. Performance in an academic setting is determined by factors relating to the opportunity, willingness, and capacity to perform well (Poropat, 2009; Van Gog, Ericsson, Rikers, & Paas, 2005). Willingness to perform portrays a stimulation to act which is usually triggered by an incentive and reflects personality (Blumberg & Pringle, 1982; Poropat, 2009). Furthermore, factors associated with willingness to perform well such as initiative, sporting activities, motivation and attitudes to study have also been shown to predict academic performance (Pintrich & De Groot, 1990; Trudeau & Shephard, 2008; Willingham, Pollack, & Lewis, 2002). Therefore it is logical to expect personality and affective state to be correlated with academic performance. Performance has been largely used by researchers to determine learning outcomes (G. Corbett, Hippisley, Brown, & Marriott, 2001; Sweller, Van Merrienboer, & Paas, 1998). Most adaptive ITSs have used learner performance as a core characteristic for adaptation. We have previously investigated performance and personality on adaptive exercise selection and performance has proved to be a strong determinant with exercise difficulty being adapted to past performance (Okpo, Dennis, Smith, et al., 2016; Okpo, Masthoff, Dennis, Beacham, & Ciocarlan, 2017). From the results of the meta-analysis by Poropat (2009) on personality and academic performance, it has also been shown that personality is definitely associated with performance (in particular, agreeableness, conscientiousness and openness).

Support used: The use the learner has made of available support in doing past exercises (for example, use of hints) should also be considered by an adaptive exercise selection algorithm, as it impacts on mental effort used, and learning achieved.

2.3. Interface Model

The *Interface Model* is often seen as part of the adaptation model, though some consider it separately. It manages the communication between learner and computer, and the presentation of the instructional content. Exercise selections will need to be presented to learners, and we envisage that adaptation of this presentation will be required, hence the inclusion of an adaptive exercise presentation algorithm in Figure 1. However, as this is not the current focus of our research, we do not elaborate further here.

2.4. Adaptation Model

The *Adaptation model* describes the tutoring strategies used to achieve adaptation. It controls the workings of the adaptive system. This paper focuses on adaptive exercise selection. Exercise selection (also referred to as task selection) is the choosing of exercises for learning and assessment. An instructional exercise is seen as engaging in a learning activity so as to develop specific skills (Stein & Lane, 1996). The International Commission on Mathematical Instruction (ICMI) (Kilpatrick, 2014) defines learning tasks as tools that bridge the gap between tutoring and learning. The ICMI also regards exercise selection to be of paramount importance in tutoring as it plays a major role in determining the next level in learning as well as the quality of the learning outcomes. There are several reports of a strong connection between the tasks selected for a learner and the learner's overall academic performance (Corbalan et al., 2008). Effective exercise selection (Rosenberg, 1986) is crucial as it culminates in the final assessment of a learners performance in any particular domain. Furthermore, as performance is regarded as the effectiveness in accomplishing an academic exercise, suitable exercise selection is a process that leads to the attainment of high grades by a learner, and grades are the dominant measures of academic performance as attested by their frequent use in criterion variable in several studies (Kuncel, Credé, & Thomas, 2005; Poropat, 2009).

2.4.1. Approaches to Exercise Selection

Different approaches to exercise selection have been used across various domains for different kinds of training purposes (Salden, Paas, & Van Merriënboer, 2006).

Static Part-task Selection Approach. This approach considers a pre-set order of learning tasks as well as the complexity of the learning task which are determined before the training. Learners are expected to start with a part-task which can include fractions, segments, simplifications or prerequisites of the whole task and practice increasingly larger parts until mastery of the whole task is achieved. In this approach, the determination of the different parts in the tutoring can be: (1) *Backward chaining* where earlier components of the tasks are introduced after the last component has been practised; (2) *Forward chaining* where task components are added sequentially from first to last (Salden, Paas, & Van Merriënboer, 2006).

Hierarchical approach. The hierarchical approach is based on the demand for prerequisites such that the basic skills must be learned before a more complex skill can be learned (Gagne, 1968; Reigeluth & Carr-Chellman, 1999).

Static whole Task Selection Approach. In this approach, it is assumed that all learning tasks require learning skills that the learner should have acquired for onward application after the training process. Also, there is the claim (Salden, Paas, &

Van Merriënboer, 2006; Taminiau et al., 2013) that training should be adjusted to the experience of the learner and task complexity should increase alongside with the experience of the learner.

Dynamic Part-Task Selection Approach. As tutoring started becoming computer-based, it was realised that it was possible to make changes to the order of the exercises as well as the complexity of the exercises during the tutoring process. As a reasonable level of skill is obtained by the learner, additional tasks can be presented to test the expertise of the learner.

Dynamic Whole-Task Selection Approach. Computer-based tutoring paved the way for the use of dynamic whole task approaches in complex tutoring. This is because adaptation to the specific needs of individual learners is made possible through this approach even while the training is in progress. This dynamism which takes place while the learner interacts with its environment is typically what ITS aims to achieve (Vandewaetere, Desmet, & Clarebout, 2011). The dynamic whole task selection approach is therefore seen as the modern approach to task selection, and is adopted in this paper.

2.4.2. Adaptive Exercise Selection

In the area of exercise selection, the focus has been on the design of intelligent tutors that select learning tasks based on the learner's past performance, available learning support and recently, cognitive load (e.g. Camp et al., 2001; Corbalan et al., 2008, 2011; Kostons et al., 2010; Salden, Paas, & Van Merriënboer, 2006; Stephens et al., 2013). Van Gog et al. (2005) explores how the activities and principles in expert performance research can be used to design instructional formats based on cognitive load theory for skills mastery. In this work, they showed that learning tasks can be adaptively selected on the basis of online assessment of a learner's expertise level.

Exercise selection is sometimes regarded as a self-directed learning skill which enables learners to select a task themselves that best fits their learning needs as provided by self-assessment (Taminiau et al., 2015). In their study, a learner needs to determine if the subsequent task should contain less support, equal support, or more support, or if it should be less difficult, equally difficult or more difficult than the previous task. However, other studies on exercise selection have provided empirical evidence that learners often do not have sufficiently developed self-directed learning skills to select suitable tasks (Kicken, Brand-Gruwel, & van Merriënboer, 2008). The studies in this paper will inform an adaptive exercise selection algorithm.

The selection of academic exercises for learners can be influenced by several factors which are also determinants of academic performance. These factors can be gender, age, self-esteem, other personality traits, academic confidence, and experience in the learning domain or with the learning content, amount of support for the task, cognitive load, feedback, interests, language or intent. One or a combination of these factors can influence academic performance (Sun, Tsai, Finger, Chen, & Yeh, 2008). We will investigate adapting exercise selection to learner personality and cognitive efficiency.

There is little research on exercise selection and personality. Existing studies have mainly focussed on locus of control. Studies have shown that learners with an internal locus of control have greater motivation in their commitment to completing tasks which are novel, challenging and daring (Devin, Ghahramanlou, Fooladian, & Zohoorian, 2012; Hoffman et al., 2016; Meeks, Hendrick, & Hendrick, 1998). They also tend to believe in their personal abilities to achieve goals and obtain domain mastery rather easily. On the contrary, due to the personal beliefs of learners with external locus of

Table 2. Exercises Table. All exercises used 2 baskets. Six exercises were used of each type.

Type	No. of Balls	Ball Distribution	Basket total	Example exercise	Validated Difficulty
1	3	1 – 2	<10	④ ⑨ ⑤	1
2	4	1 – 3	<10	③ ① ⑤ ⑨	2
3	4	2 – 2	<10	⑥ ⑤ ③ ④	2
4	4	2 – 2	≥10	⑦ ⑤ ④ ⑥	3
5	5	1 – 4	<10	① ① ⑨ ② ⑤	3
6	5	2 – 3	<10	⑥ ② ③ ④ ③	4
7	5	2 – 3	≥10	② ⑧ ⑥ ⑨ ③	5

control (that they are not responsible for the outcome of their learning) they tend to lack effectiveness in different learning situations (Devin et al., 2012). Similarly, Krause and Broderick (2006) found that for predicting outcomes of learning, it is important to consider locus of control. With regard to exercise selection, Holloway (1978) showed a significant locus of control and exercise selection interaction for learners with high ability which points to the consideration of personal control in selection of instruction in tutoring.

Adaptive exercise selection may need to consider other personality traits, such as the FFM traits. For example, one could conjecture that a learner's Openness to Experience may impact on a learner's willingness to try new exercises, and a learner's Emotional Stability on the level of support and scaffolding needed. Conscientiousness may also be relevant, as perhaps conscientious learners can be given more exercise repetitions and conscientiousness may determine the mental effort put in. Extroversion may affect the exercise style (group vs individual), as investigated by Alhathli et al. (2016).

In this paper, building on our research (Okpo, Masthoff, Dennis, Beacham, & Cio-carlan, 2017), we will investigate the link between exercise selection and self-esteem. Adaptive exercise selection may aim to boost learner's confidence, particularly for those with low-self esteem. This paper expands this research by investigating the effect of additional levels of performance and producing an algorithm based on this new data.

3. Development of Study Materials

In this section, we describe the development and validation of the materials used in later studies. These materials include a set of exercises with various difficulties, the stories used to express the self-esteem of the learner, and the sentences used to describe the effort they invested in completing an exercise. These will be used to investigate the selected difficulty of the 'next exercise' for a learner, based on their performance, effort, and personality.

3.1. Exercise Difficulty Study Validation

To investigate the effect of mental effort and personality on the difficulty of exercises, we first needed a set of exercises with validated difficulty levels for participants to select from. In this paper, we have used a simple addition exercise which asks learners

Figure 2. Interface for Difficulty Validation Study

to place different weighted balls into a set number of baskets so that they weighed the same. For example, a learner might be given three balls: (2), (3) and (5) and asked to place them into two baskets. This is a very easy exercise as the heaviest ball is also the sum of the remaining two. Exercises can be made more challenging in many ways. For example, requiring more than one ball in both baskets: (2) (3) in one basket and (4) (1) in the other; increasing the basket total so that more balls are needed: (5) (4) (7) (9) in one and (3) (8) (2) (4) (6) in the other; and by increasing the number of baskets. With this in mind, we wanted to create a systematic way of increasing difficulty using these methods. To achieve this, we performed a validation study.

3.1.1. Study Design

Participants were recruited via Amazon's Mechanical Turk (MT, 2012), a crowd-sourcing tool. Participants required an acceptance rate of 90% to ensure good quality of responses and had to pass a Cloze test (Taylor, 1953) for English fluency to ensure they had enough literacy skills to understand the language used for the study. 155 participants took part in the study (74 females and 81 males).

Six pairs of two exercises with different estimated difficulty levels were shown to the participants. These exercises were generated using the rules described in Table 2. As previously discussed, we varied the complexity of the exercises by adjusting the number of balls required to be placed in each basket to solve the exercise, and by increasing the basket total.

We had six variants of the study in a between-subjects design, each investigating the difference in difficulty between two types of exercises. Each variant had twelve exercises (six of each type) for participants to solve, shown in pairs (one exercise of each type), where one exercise was expected to be easier than the other. Participants then rated which exercise they thought was easier on a scale as shown in Figure 2. Within each pair, the order was randomized so that the exercise we expected to be easier could appear on the left or right.

Table 3. Results of Difficulty Validation. Significance (p) values obtained by z-test of grand mean compared to 0.

Study	Exercise Types	Num. Participants	Mean	Std. Dev	p
1	1 & 2	25	0.74	0.944	0.000
2	2 & 3	25	-0.13	0.808	0.978
3	3 & 4	25	0.20	0.993	0.006
4	4 & 5	25	-0.29	1.006	0.999
5	4 & 6	25	0.43	1.226	0.000
6	6 & 7	25	0.50	1.184	0.000

3.1.2. Results

To calculate a relative difficulty score between two exercises of different types, we transformed the scale shown in Figure 2 into numbers. If the exercise with the harder expected difficulty was deemed slightly harder this was scored as 1, harder as 2 and much harder as 3 (and corresponding negative scores for the reverse). If the exercises were deemed of the same difficulty a score of 0 was used. The score was averaged over the six comparisons of the two exercise types each participant did.

From our results in Table 3, the Z-test of the grand mean scores compared to 0 shows that exercise types 1 and 2 had a significant difference and therefore type 2 is more difficult than type 1. Exercise types 2 and 3 showed no significant difference. Exercise type 3 was then tested with exercise type 4 which gave a significant result showing that type 4 was more difficult than type 3 (or 2). Testing exercise type 4 against type 5 yielded no significant difference indicating that types 4 and 5 were of the same difficulty. We then tested exercise type 4 with type 6 and this returned a significant difference in difficulty with type 6 being more difficult than type 4 (or 5). Finally, exercise type 6 was tested with type 7 and the result showed a high significant difference proving that type 7 was more difficult than type 6. Consequently, we have been able to derive 5 levels of difficulty of exercises (see Table 2).

3.2. Self-Esteem Story Validation

This section describes the creation and validation of stories which express learner self-esteem (SE) at polarized levels (high and low), following a similar approach to Dennis et al. (2012).

3.2.1. Story Development

To construct the stories, we used the well-established State Self-Esteem scale (SSES) (Heatherton & Polivy, 1991). The SSES consists of 20 items that measure momentary fluctuations in self-esteem. For each story, we changed a selection of the questionnaire items into the third person, inverting them where necessary. In trying to make the story real, we linked it with a character, a student called *Nancy*. The resulting stories are shown in Table 4.

3.2.2. Story Validation

40 participants saw one of the two stories in a between-subjects design (following a similar approach to Dennis et al. (2012)). Participants were crowd-sourced on Amazon's Mechanical Turk (MT, 2012). They were adults based in the US, had to pass an English fluency test and have an approval rating of 90%. To validate that the stories

Table 4. Stories depicting High and Low Self Esteem

SE Level	Story
High	Nancy is a learner who is confident about her abilities. She is satisfied about the way she looks and feels good about herself. She thinks she is as smart as others and believes that others admire and respect her. She feels that she has a good understanding of things.
Low	Nancy is a learner who worries about the impression she makes and whether she is regarded as a success or a failure. She feels like she is not doing well and she believes she cannot understand the things she reads. Nancy thinks she is unattractive and is displeased with herself. She feels inferior to others.

Table 5. Mental Effort Statements Mapped to Effort Rating. All statements ended with 'effort'. Statements in bold were used in the follow-on studies.

Statement	1 (none)	2	3	4	5	6	7	8	9	10 (max)	Average Rating
no	95%		5%								1.10
minimal	24%	**52%**	14%	10%							2.10
little		24%	**71%**				5%				2.95
token		35%	10%	10%	15%	20%		5%		5%	4.25
some		5%	14%	52%		24%	5%				4.38
average				10%	71%	10%	10%				5.19
moderate				14%	**48%**	24%	10%			5%	5.52
quite a lot of		5%			10%	14%	33%	33%		5%	6.90
considerable					10%	24%	38%	14%	14%		7.00
substantial					10%	25%	30%	25%	10%		7.00
much			5%			14%	**52%**	24%	5%	0%	7.00
very much						10%	10%	48%	29%	5%	8.10
ultimate						10%	5%	0%	10%	76%	9.38
maximum						5%	5%		10%	81%	9.57
all possible							5%	5%	10%	**80%**	9.65

conveyed self-esteem at the desired level, participants were asked to rate the self-esteem of Nancy using the Rosenberg self-esteem questionnaire (Rosenberg, Schooler, Schoenbach, & Rosenberg, 1995), which uses different terms to measure the same concept.

A between subjects T-test was performed on the self-esteem score between the high and low self-esteem stories, which was significant at $T(38) = 13.93$, $p < 0.001$. The mean self-esteem score for high was 24.6 ± 3.68 SD and the mean for low was 8.0 ± 3.87 SD. The Rosenberg scale ranges from 0-30. Scores between 15 and 25 are within the normal range, while scores below 15 suggests low self-esteem (Rosenberg et al., 1995). The level of self-esteem for the low story is lower than the normal range. The level of self-esteem for the high story is at the top end of the normal range.

Given the large difference in the self-esteem levels between the two stories, we believe the stories are suitable for future studies to express high and low learner self-esteem.

3.3. Mental Effort Statement Validation

3.3.1. Statement Generation

Many statements can be used to describe the amount of mental effort that has been invested by a learner in carrying out a task. Unfortunately, we have not found a list that clearly defines varying levels of mental effort. To have a clear definition of invested mental effort, we wanted to map mental effort statements to numbers indicating the

effort used, so that we could use a selection of statements in our studies. In a brainstorming session, three lecturers and a research student came up with 15 invested mental effort statements (see statements in Table 5).

3.3.2. Statement Validation

26 participants (staff and students of the University of Aberdeen) completed the Mapping Numbers to Mental Effort Statements survey using an on-line survey tool (Survey Monkey). The data from 21 participants were used (16 males, 4 females, 1 undisclosed). The other 5 participants were excluded due to the low quality of their responses (providing the same response for almost all statements, or not mapping 'no effort' to 1 as was indicated on the scale).

Participants read the mental effort statements and mapped them to numbers from 1 to 10 with 1 representing no effort and 10 representing maximum effort. The order of the mental effort statements was randomized for each participant.

Table 5 shows the percentage of participants who mapped a statement to a particular number. Some statements (e.g. 'token effort') showed little agreement between participants, whilst others showed better agreement. We decided to use five statements (shown in bold) for the main studies (see section 4.3 for the selection rationale).

4. Investigating the Impact of personality and Mental Effort on Exercise Selection: Studies' Design

Above we presented studies on the validation of a set of exercises with varying difficulty, a set of validated mental effort statements and a set of validated personality trait stories for self-esteem. Using these materials, we performed three studies, one for each of our three performance conditions: "just passed", "just failed", and "performed well". We used the same design for all conditions.

4.1. Methodology

We used the User-as-Wizard (Masthoff, 2006) methodology for our studies where the role of the system is played by participants in selecting exercises for a learner to do next. These studies build on previous research investigating the impact of personality and performance in selecting the next exercise for learners (Okpo, Dennis, Smith, et al., 2016) by including invested mental effort and using validated difficulty levels of exercises. Following the methodology used by (Dennis et al., 2016), in this paper, we use crowd-sourcing to inspire the algorithm. The resulting algorithm will later be validated with teachers, followed by an investigation of the actual impact on learners.

4.2. Variables

The dependent variable for the studies is the difficulty level of the exercise selected for Nancy to do next. Participants could choose between 5 difficulty levels (level 1 to 5). They were told that Nancy had done 10 exercises of difficulty level 2 before. So, participants could select exercises of the same difficulty (level 2), or an easier difficulty (level 1) or of varying degrees of more difficulty (levels 3-5).

The independent variables used for the studies are: learner self-esteem (*high* and

Now that you have completed some exercises, read the story about Nancy below and recommend an exercise for her to do next.

Meet Nancy

Nancy is a learner who worries about the impression she makes and whether she is regarded as a success or a failure. She feels like she is not doing well and believes she cannot understand the things she reads. Nancy thinks she is unattractive and is displeased with herself. She feels inferior to others.

Nancy completed a set of mathematics exercises of difficulty level 2. Each exercise required the addition of 4 balls representing 1 digit numbers. The totals in the baskets are 1 digit numbers.(For example: ④, ①, ②, and ⑦).

Ten exercises were given to Nancy and she **just passed.** Nancy put **all possible effort** into solving the exercises.

Assuming you were Nancy's Mathematics teacher, which next exercise would you give to Nancy from the list of exercises below?

Select the exercise that you would give to Nancy by clicking on the button on the right of your chosen exercise.

Exercise	Example	Difficulty	Select
Addition of 3 balls representing 1 digit numbers. The totals in the baskets are 1 digit numbers. For example: **9, 2 and 7**. These are an easier type of exercise than the ones she did before.	9 2 7	**Difficulty Level 1**	○
Addition of 4 balls representing 1 digit numbers. The totals in the baskets are 1 digit numbers. For example: **3, 1, 5 and 9** These are the same type of exercise that she did before.	3 1 5 9	**Difficulty Level 2**	○
Addition of 4 balls representing 1 digit numbers. The totals in the baskets are 2 digit numbers. For example: **7, 5, 4 and 6** These exercises are more difficult than those she did before.	7 5 4 6	**Difficulty Level 3**	○
Addition of 5 balls representing 1 digit numbers. The totals in the baskets are 1 digit numbers. For example: **6, 2, 3, 4 and 3** These exercises are considerably more difficult than those she did before.	6 2 3 4 3	**Difficulty Level 4**	○
Addition of 5 balls representing 1 digit numbers. The totals in the baskets are 2 digit numbers. For example: **2, 8, 6, 9 and 3** These exercises are the most difficult of all exercises.	2 8 6 9 3	**Difficulty Level 5**	○

Figure 3. Screenshot of the exercise selection stage of the study for condition "low self-esteem", "just passed", and "all possible effort"

low) and their invested mental effort in solving the exercises (*minimal effort, little effort, moderate effort, much effort,* and *all possible effort*).

Participants were also told about Nancy's performance, and this varied depending on the experiment: she *just passed* in Experiment 1, *just failed* in Experiment 2, and *performed well* in Experiment 3.

4.3. Materials

Validated stories conveyed the self-esteem of a fictional learner 'Nancy' (see Table 4). A sentence was added to indicate her past performance, and one of the validated mental effort statements was shown from the bold items in Table 5. The statements were selected to ensure a good spread of difficulties, and had good inter-rater agreement. However, we decided to exclude "no effort" and "maximum effort" as they were used in the explanation of the scale that participants saw in the validation experiment, and we excluded "average effort" as this could be affected by the learning domain. In addition, a set of exercises with validated difficulty levels (see Table 2) were presented to participants to select the one Nancy should do next.

4.4. Procedure

Participants began by completing five short exercises just like the ones that the learners would do so they gained an understanding of the different difficulty levels. The order of exercises presented to participants was from easiest to most difficult. In a between-subjects design, participants were then asked to select the exercise Nancy should do next, given her self-esteem, past performance and invested mental effort, shown in Figure 3. Participants were informed that it was their opinion that counted and as such there were no right or wrong answers.

4.5. Hypotheses

- **H1**: The mean difficulty level selected by participants will be higher for High SE learners than for Low SE learners.
- **H2**: The mean difficulty level selected by participants will be higher for learners with a lower mental effort than higher mental effort.
- **H3**: Participants will select a different difficulty level for the exercise depending on the combination of SE level and mental effort.
- **H4**: When combining the results of the three studies, participants will select a more challenging exercise for learners who performed better than for learners who performed worse.

4.6. Participant Recruitment

All studies were administered through crowd-sourcing on Amazon's Mechanical Turk (MTurk) (MT, 2012). To be eligible to be part of a study, participants had to pass an English Cloze test. Participants were also required to successfully complete 5 short exercises (of difficulty levels 1 to 5) similar to the ones that they could select for the learners to do next. In addition, participants had to have a 90% acceptance rate meaning that 90% of the work they do on MTurk is accepted by other requesters as being of good quality.

Table 6. Exercise Selection Study Results for Performance Level Just Passed

| | Selected difficulty level (% of participants) | | | | | | | | | |
| | 1 | | 2 | | 3 | | 4 | | 5 | |
Effort	High SE	Low SE	High SE	Low SE	High SE	Low SE	High SE	Low SE	High SE	Low SE
minimal	0	5	0	25	40	35	25	25	35	10
little	5	10	5	25	43	30	29	15	19	20
moderate	5	10	10	30	60	50	20	10	5	0
much	0	20	10	15	80	40	10	15	0	10
all possible	0	20	15	20	60	50	20	5	5	5
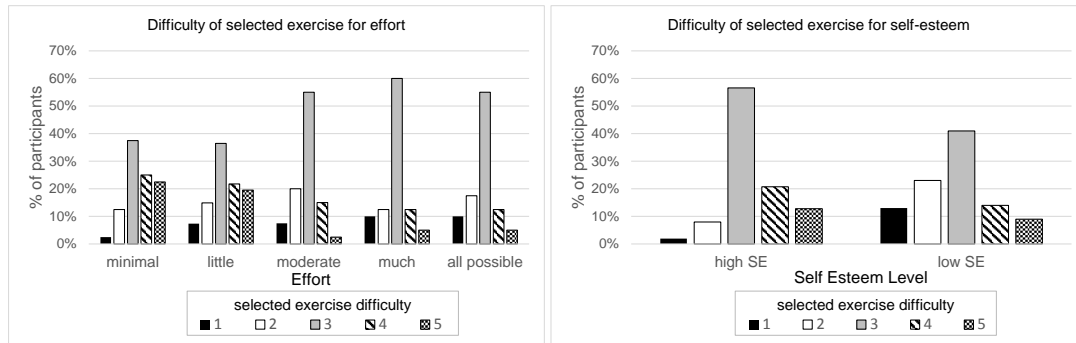
Figure 4. Chosen exercise difficulty for effort (left) and self-esteem (right) for performance level Just Passed.

5. Experiment 1: Performance Level Just Passed

This experiment investigates the impact of self-esteem and effort level on exercise selection for learners who have *just passed*.

5.1. Participants

201 participants successfully completed the study (53.2% males, 46.8% females; 12.9% aged 18 - 25, 56.7% aged 26 - 40, 29.4% aged 41 - 65, 0.5% aged 65 and over, and 0.5% undisclosed; 13.4% were students, 4.5% were teachers and 82.1% other).

5.2. Results

Table 6 shows the results for each condition. We ran a 2-way ANOVA of the independent variables *self-esteem* × *effort* for *difficulty*. This was significant for both *effort* ($F(4, 200) = 4.12$, $p < 0.005$) and *self-esteem* ($F(1, 200) = 14.04$, $p < 0.001$), however, the interaction of *effort* × *self-esteem* was not significant.

Figure 4 shows the results overall for effort and self-esteem. Looking at effort alone, it can be seen that a difficulty of 3 (*slightly harder than before*) is the most popular choice for all levels of effort. However, when effort is minimal or little, a higher percentage of participants recommend a higher level of difficulty than this (more choose a difficulty of 4 or 5 than of 3). H2 is thus confirmed, participants did choose an exercise of higher difficulty for learners who required little effort to complete the exercise.

For self-esteem, Figure 4 shows that although an exercise of slightly harder difficulty (level 3) remains the most popular choice in both conditions, for low self-esteem, participants pick an exercise of lower or the same difficulty more often than in the high condition. Thus there is support for hypothesis H1.

With respect to hypothesis H3, we have evidence that both Self-Esteem and mental

Table 7. Exercise Selection Study Results for Performance Level Just Failed

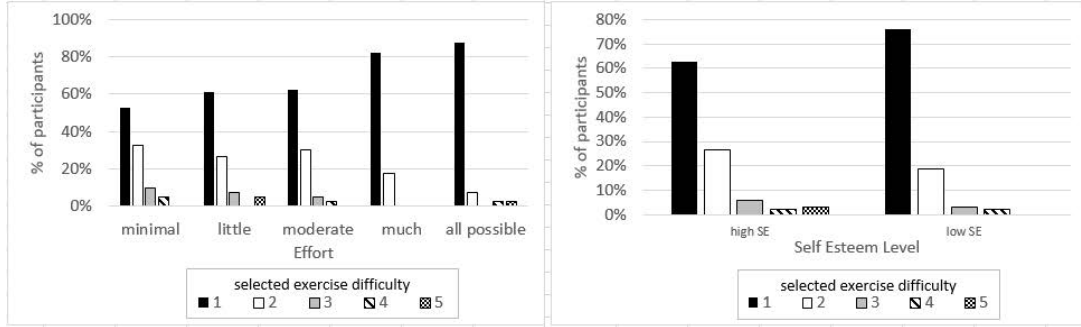| | Selected difficulty level (% of participants) | | | | | | | | | |
| | 1 | | 2 | | 3 | | 4 | | 5 | |
Effort	High SE	Low SE	High SE	Low SE	High SE	Low SE	High SE	Low SE	High SE	Low SE
minimal	35	70	45	20	15	5	5	5	0	0
little	52	70	29	25	10	5	0	0	10	0
moderate	60	65	35	25	5	5	0	5	0	0
much	85	80	15	20	0	0	0	0	0	0
all possible	80	95	10	5	0	0	5	0	5	0

Figure 5. Chosen exercise difficulty for effort (left) and self-esteem (right) for performance level Just Failed.

effort mattered to participants when selecting exercises overall, but as there is no interaction effect, we do not have strong enough evidence to factor in both mental effort and self-esteem at the same time, meaning this is not well supported.

6. Experiment 2: Performance Level Just Failed

This experiment investigates the impact of self-esteem and effort level on exercise selection for learners who have *just failed*.

6.1. Participants

202 participants successfully completed the study (55.9% males, 44.1% females; 14.9% aged 18 - 25, 61.4% aged 26 - 40, 22.8% aged 41 - 65 and 1.0% aged over 65; 18.8% were students, 5% were teachers and 76.2% other).

6.2. Results

Table 7 shows the results for each condition. We ran a 2-way ANOVA of the independent variables *self-esteem* × *effort* for *difficulty*. This was significant for both *effort* ($F(4, 192) = 3.24$, $p < 0.05$) and *self-esteem* ($F(1, 192) = 5.42$, $p < 0.05$), however again, the interaction of *effort* × *self-esteem* was not significant.

Figure 5 shows the results overall for effort and self-esteem. Looking at effort alone, it can be seen that a difficulty of 1 (*slightly easier than before*) is the most popular choice for all levels of effort, though the proportion of participants selecting an exercise with the same difficulty as the learner's previous exercise increases with decreasing levels of effort. In particular for minimal effort, particants' opinion was about evenly split on whether to decrease the difficulty level or not (keep it the same or even increase

it). In fact, the average for minimal effort is difficulty level 2, whilst the average for all possibile effort is difficulty level 1. The significant main effect of effort supports hypothesis H2.

For self-esteem, Figure 5 shows that although an exercise of slightly easier difficulty (level 1) remains the most popular choice in both conditions, for low self-esteem, participants pick an exercise of lower difficulty more often than in the high condition. And of course similarly, for high self-esteem, participants pick an exercise of the same or higher difficulty more often than in the low condition. The mean difficulty level for low self-esteem is 1, whilst the mean difficulty level for high self-esteem is 2. Thus, the significant main effect of self-esteem supports hypothesis H1.

For both effort and self-esteem, a limitation of this Experiment is that participants did not have the option to select a considerably easier exercise, given the starting level of difficulty 2 (which was done to make the results comparable to those of Experiment 1 to enable the analysis presented in Section 8). So, it is possible that some participants would have selected considerably easier exercises for lower levels of effort and low self-esteem if they had had the opportunity to do so.

With respect to hypothesis H3, similar to Experiment 1, we have evidence that both Self-Esteem and mental effort mattered to participants when selecting exercises overall, but as there is no interaction effect, we do not have strong enough evidence to factor in both mental effort and self-esteem at the same time, meaning this is not well supported.

7. Experiment 3: Performance Level Performed Well

This experiment investigates the impact of self-esteem and effort level on exercise selection for learners who have *performed well*.

7.1. Participants

206 participants successfully completed the study (50.5% males, 48.1% females and 1.5% undisclosed; 15% aged 18–25, 57.3% aged 26–40, 26.7% aged 41–65, 0.5% aged over 65 and 0.5% undisclosed; 15% were students, 6.8% were teachers and 78.2% other).

7.2. Results

Table 8 shows the results for each condition. We ran a 2-way ANOVA of the independent variables *self-esteem* × *effort* for *difficulty*. This was significant for *self-esteem* ($F(1, 196) = 8.49$, $p < 0.005$), but not for *effort* ($F(4, 196) = 0.268$, $p = 0.9$). The interaction of *effort* × *self-esteem* was also not significant.

Figure 6 shows the results overall for effort and self-esteem. Looking at effort alone, it can be seen that a difficulty of 3 (*slightly harder than before*) was the most popular choice. Given the lack of a main effect of effort, there is no clear evidence to support hypothesis H2, though it can be seen that participants choose difficulty 2 (*same difficulty as before*) more often for learners who put in "much effort" or "all possible effort" than for learners who put in less effort.

For self-esteem, Figure 6 shows that whilst a difficulty of 3 (*slightly harder than before*) remains the most popular choice in both conditions, for high self-esteem, a similar proportion of participants wants a higher difficulty of 4 or 5 as want one of 3.

Table 8. Exercise Selection Study Results for Performance Level Performed Well

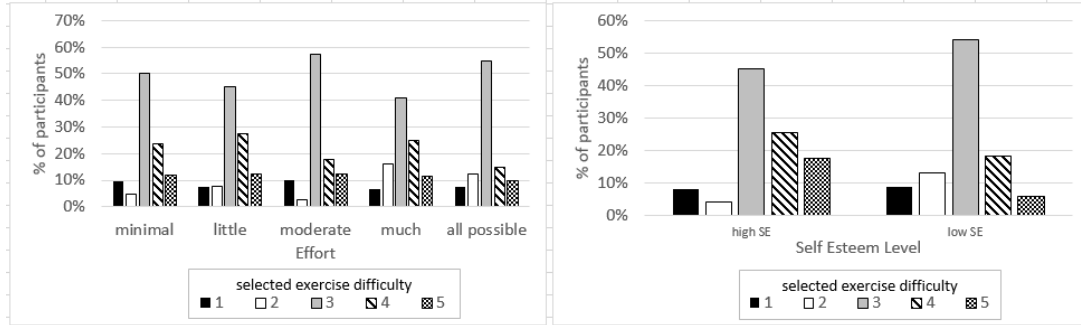| | Selected difficulty level (% of participants) | | | | | | | | | |
| | 1 | | 2 | | 3 | | 4 | | 5 | |
Effort	High SE	Low SE	High SE	Low SE	High SE	Low SE	High SE	Low SE	High SE	Low SE
minimal	10	9	0	9	50	50	25	23	15	9
little	5	10	5	10	45	45	25	30	20	5
moderate	5	15	5	0	45	70	30	5	15	10
much	5	9	0	32	41	41	36	14	18	5
all possible	15	0	10	15	45	65	10	20	20	0

Figure 6. Chosen exercise difficulty for effort (left) and self-esteem (right) for performance level Performed Well.

On the other hand, for low-self esteem, there is a clear preference for a difficulty of 3. Thus there is support for hypothesis H1.

With respect to hypothesis H3, given the lack of an interaction effect, we have not enough evidence to factor in both self-esteem and effort at the same time, meaning H3 is not supported.

8. The impact of performance: Combining the three studies

The three studies reported above each individually kept learner performance static whilst varying effort and personality. We combined the data of the three studies to analyse the impact of performance, and possible interaction effects between performance on the one hand and effort and personality on the other. Figures 7, 8, and 9 show the combined data for self-esteem, effort and performance respectively.

We ran a 3-way ANOVA of the independent variables *performance × self-esteem*

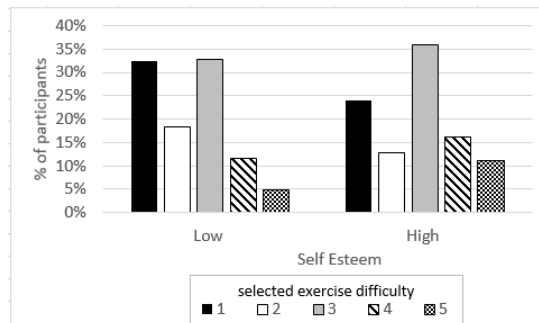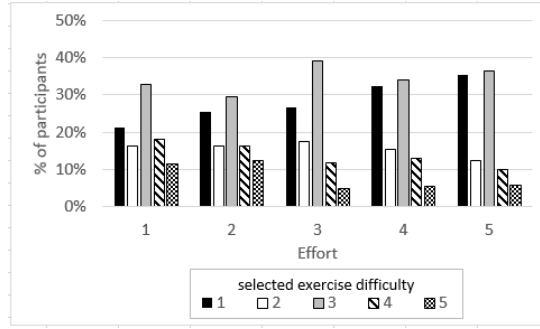**Figure 7.** Difficulty selected for high and low self-esteem across conditions.

21

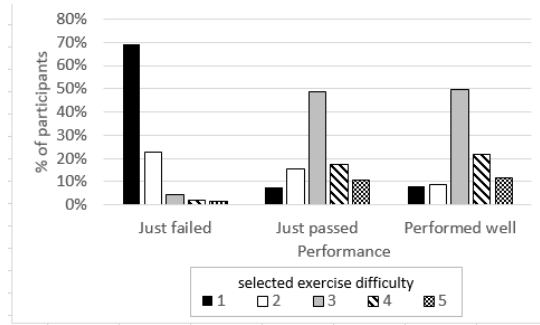**Figure 8.** Difficulty selected for effort across conditions.

Figure 9. Difficulty selected for performance across conditions.

$\times$ *effort* for *difficulty*. This was significant for *performance* ($F(2, 579) = 229.5$, $p < 0.001$), *effort* ($F(4, 579) = 5.19$, $p < 0.001$) and *self-esteem* ($F(1, 579) = 27.38$, $p < 0.001$). This supports Hypotheses H1, H2 and H4. There were no significant interaction effects between any of the independent variables, so no evidence was found to support H3.

9. Regression Analysis and Resulting Algorithm

Using the combined data of the three studies, a cumulative odds ordinal logistic regression with proportional odds was run to predict difficulty level based on self-esteem, effort and performance[1]. The proportional odds assumption was only partially met: independent binominal logistic regression analyses showed that the assumption is tenable for effort and self-esteem, but not for performance. This means that the results regarding performance below need to be treated with caution.

The final model statistically significantly predicted the difficulty level over and above the intercept-only model, $\chi^2(3)=244.990$, p <.0005[2]. The odds ratio of selecting a higher difficulty level for learners with high self-esteem versus low self-esteem is 1.884 (95% CI, 1.397 to 2.540), a statistically significant effect, Wald $\chi^2(1) = 17.271$, p <.0005. A decrease in effort was associated with selecting a higher difficulty level with

---

[1]Self-esteem was used as a factor. Effort level and performance were used as ordinal co-variates, with effort level coded 1-5 for minimal effort till all possible effort and performance coded 1-3 for performed well to just failed

[2]Given many cells were sparse with zero frequencies in 19.3% of cells, the Deviance goodness of fit is not appropriate to consider.

an odds ratio of 1.261 (95% CI, 1.135 to 1.403), Wald $\chi^2(1) = 18.463$, p $<.0005$. An increase in performance was associated with selecting a higher difficulty level with an odds ratio of 4.367 (95% CI, 3.534 to 5.405), Wald $\chi^2(1) = 185.74$, p $<.0005$. This provides evidence that performance has the most impact on difficulty level selection, followed by self-esteem, followed by effort. This finding remains true when considering the independent binary regressions, with the odd ratio for performance always higher than that of the other variables.

The model provides coefficients to calculate a value, as well as thresholds to compare the calculated value against to produce cumulative odds for difficulty levels.

The model's coefficients result in the following formulae to calculate Value:

- $-1.475 \times Performance - 0.232 \times Effort + 0.633$, if $Self\text{-}esteem = $ high
- $-1.475 \times Performance - 0.232 \times Effort$,   if $Self\text{-}esteem = $ low

The thresholds lead to the following formulae to calculate the natural logarithm of the cumulative odds:

- $\ln(\text{Odds}(difficulty \leq 1)) = -4.577 - Value$
- $\ln(\text{Odds}(difficulty \leq 2)) = -3.573 - Value$
- $\ln(\text{Odds}(difficulty \leq 3)) = -1.588 - Value$
- $\ln(\text{Odds}(difficulty \leq 4)) = -0.315 - Value$

Using these formulae, for each combination of performance, self-esteem, and effort we calculated:

- Value, see Table 9
- Odds($difficulty \leq d$), for all difficulty levels d
- Probability P($difficulty \leq d$), for all difficulty levels d
- P($difficulty = d$) for all difficulty levels d, using that
  P($difficulty \leq 1$)=P($difficulty = 1$) and
  P($difficulty = d + 1$) = P($difficulty \leq d+1$) $-$ P($difficulty \leq d$)
- Median difficulty m such that
  P($difficulty \leq m$) $\geq 0.5 \wedge$ P($difficulty \geq m$) ≥ 0.5

Table 9 shows the calculated values for all our combinations of Performance, Self-Esteem, and Effort, and how these values map onto median difficulty levels. The predicted median difficulty levels were used to produce our algorithm, presented in Algorithm 1.

10. Conclusions

This paper investigated the adaptation of exercise selection (difficulty level) to learner personality (self-esteem) and cognitive efficiency (performance and mental effort). Three studies were conducted, with over 600 participants in total, to investigate how people select exercise difficulty for a learner with a certain level of self-esteem (*high* or *low*), performance (*just passing*, *just failing*, *performed well*), and mental effort (5 levels). The results provide evidence that people take into account self-esteem and mental effort for exercise selection in addition to performance. They also provide an indication of how people think exercises can be selected to achieve a better learning experience. Based on the data analysis, an algorithm has been constructed to perform the adaptation.

Table 9. Model predictions

Performance	Self-esteem	Effort	Calculated Value	Median difficulty
Well	High	1 Minimal	-1.074	
		2 Little	-1.306	4
		3 Moderate	-1.538	
		4 Much	-1.770	
		5 All possible	-2.002	
	Low	1 Minimal	-1.707	
		2 Little	-1.939	
		3 Moderate	-2.171	
		4 Much	-2.403	
		5 All possible	-2.635	
Just passed	High	1 Minimal	-2.549	3
		2 Little	-2.781	
		3 Moderate	-3.013	
		4 Much	-3.245	
		5 All possible	-3.477	
	Low	1 Minimal	-3.182	
		2 Little	-3.414	
		3 Moderate	-3.646	
		4 Much	-3.878	
		5 All possible	-4.110	2
Just failed	High	1 Minimal	-4.024	
		2 Little	-4.256	
		3 Moderate	-4.488	
		4 Much	-4.720	
		5 All possible	-4.952	
	Low	1 Minimal	-4.657	
		2 Little	-4.889	1
		3 Moderate	-5.121	
		4 Much	-5.353	
		5 All possible	-5.585	

Algorithm 1: Algorithm for Exercise Difficulty Selection

Input : *self esteem* the learner's level of self esteem; *performance* the learner's performance on the previous exercise; *effort* the learner's effort level on the previous exercise; *old difficulty* the difficulty level of the previous exercise (should be >1)

Output: *difficulty* the difficulty level to use for the next exercise

1 **begin**
2   **switch** *performance* **do**
3   **case** *just failed* **do**
4   **if** *self esteem = low* **then**
5          $difficulty = old\,difficulty - 1$;
6   **else**
7   **if** $effort <= 3$ **then**
8            $difficulty = old\,difficulty$;
9   **else**
10            $difficulty = old\,difficulty - 1$;
11   **end**
12   **end**
13   **end**
14   **case** *performed well* **do**
15   **if** $(self\,esteem = high) \wedge (effort <= 3)$ **then**
16          $difficulty = old\,difficulty + 2$;
17   **else**
18          $difficulty = old\,difficulty + 1$;
19   **end**
20   **end**
21   **case** *just passed* **do**
22   **if** *self esteem = high* **then**
23          $difficulty = old\,difficulty + 1$;
24   **else**
25   **if** $effort >= 3$ **then**
26            $difficulty = old\,difficulty$;
27   **else**
28            $difficulty = old\,difficulty + 1$;
29   **end**
30   **end**
31   **end**
32   **end**
33 **end**

In addition to insights into the adaptation of exercise selection, the research presented has also produced a set of exercises with validated difficulty levels, validated mental effort levels and validated personality trait stories for self-esteem which can be used for further studies.

The work presented in this paper has several limitations and options for future work. Firstly, we only investigated three levels of learner performance (*just passed*, *just failed* and *performed well*). We will extend this by also investigating *performed badly*. This will require us to change the starting difficulty level to 4, to provide participants with multiple gradations of easier exercises. However, if this additional study was run, it would not have been easily possible to compare the data with that of the studies reported in this paper due to the higher start difficulty level. Additionally, the algorithm presented is based on studies with a starting difficulty of 2, which makes the selection of an easier exercise possible (but not that of a much easier exercise). We have not investigated whether the results generalize to larger starting difficulty levels.

Secondly, we investigated adaptation to only one personality trait namely self-esteem. The effect of other personality traits can be investigated using existing *Personality trait stories* to allow an intelligent tutoring system to adapt to other facets of learner personality. Based on previous research (Dennis et al., 2016), we expect learner conscientiousness and neuroticism from the five-factor model to be relevant traits.

Finally, whilst the analysis performed on the combined data of the three studies inspired an algorithm, there are limitations on the way this algorithm was constructed. Firstly, due to the need for large numbers of participants (we used over 600 participants in the three main studies), this paper utilized crowd-sourcing for recruitment, which raises questions on how good crowd-workers (who are ordinary people) are at this task. Secondly, the method used to produce the algorithm from the data uses estimation. Whilst the resulting algorithm looks quite sensible, this clearly needs verification. Any resulting adaptations require the input of experts in the learning domain for further refinement and to verify that the adaptations are appropriate. Follow-on studies with teachers are planned to validate and/or refine the algorithm, followed by studies with learners to measure the effect on learner achievement, motivation and confidence.

References

Aleven, V., Mclaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. *Int. Journal of Artificial Intelligence in Education*, *16*(2), 101–128.

Alhathli, M., Masthoff, J., & Siddharthan, A. (2016). Exploring the impact of extroversion on the selection of learning materials. In *Pale workshop at umap*.

Alhathli, M., Masthoff, J., & Siddharthan, A. (2017). Should learning material's selection be adapted to learning style and personality? In *Adjunct publication of the 25th conference on user modeling, adaptation and personalization* (pp. 275–280). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/3099023.3099079`

Arevalillo-Herráez, M., Arnau, D., Marco-Giménez, L., González-Calero, J. A., Moreno-Picot, S., Moreno-Clari, P., ... others (2014). Providing personalized guidance in arithmetic problem solving..

Assis, M. A. d., Mello, M. F. d., Scorza, F. A., Cadrobbi, M. P., Schooedl, A. F., Silva, S. G. d., ... Arida, R. M. (2008). Evaluation of physical activity habits in patients with posttraumatic stress disorder. *Clinics*, *63*(4), 473–478.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Prentice-Hall, Inc.

Beal, C., & Lee, H. (2005). Creating a pedagogical model that uses student self reports of motivation and mood to adapt its instruction. In *Workshop on motivation and affect in educational software, in conjunction with the 12th international conference on artificial intelligence in education.*

Bloom, B. S. (1956). Taxonomy of educational objectives: The classification of educational goals .

Blumberg, M., & Pringle, C. D. (1982). The missing opportunity in organizational research: Some implications for a theory of work performance. *Academy of management Review*, *7*(4), 560–569.

Brusilovsky, P., Schwarz, E., & Weber, G. (1996). Elm-art: An intelligent tutoring system on world wide web. In *Intelligent tutoring systems* (pp. 261–269).

Camp, G., Paas, F., Rikers, R., & van Merrienboer, J. (2001). Dynamic problem selection in air traffic control training: A comparison between performance, mental effort and mental efficiency. *Comput Hum Behav*, *17*(5), 575–595.

Cheng, I., Shen, R., & Basu, A. (2008). An algorithm for automatic difficulty level estimation of multimedia mathematical test items. In *Advanced learning technologies, icalt'08* (pp. 175–179).

Conati, C. (2009). Intelligent tutoring systems: New challenges and directions. In *Ijcai* (Vol. 9, pp. 2–7).

Conlan, O., Dagger, D., & Wade, V. (2002). Towards a standards-based approach to e-learning personalization using reusable learning objects. In *Proc. of world conference on e-learning, e-learn* (pp. 15–19).

Corbalan, G., Kester, L., & van Merriënboer, J. J. (2008). Selecting learning tasks: Effects of adaptation and shared control on learning efficiency and task involvement. *CEP*, *33*(4), 733–756.

Corbalan, G., Kester, L., & Van Merrienboer, J. J. (2011). Learner-controlled selection of tasks with different surface and structural features: Effects on transfer and efficiency. *Computers in Human Behavior*, *27*(1), 76–81.

Corbett, A. T., Koedinger, K. R., & Anderson, J. R. (1997). Intelligent tutoring systems. *Handbook of human-computer interaction*, *5*, 849–874.

Corbett, G., Hippisley, A., Brown, D., & Marriott, P. (2001). A perspective from russian on a complex relation¿ 1. *Frequency and the emergence of linguistic structure*, *45*, 201.

Corno, L., & Rohrkemper, M. (1985). The intrinsic motivation to learn in classrooms. *Research on motivation in education*, *2*, 53–90.

Costa, P. T., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and individual differences*, *13*(6), 653–665.

Csikszentmihalyi, M. (1988). The flow experience and its significance for human psychology. In (p. 15-35). Cambridge, UK: Cambridge University Press.

Davidovic, A., Warren, J., & Trichina, E. (2003). Learning benefits of structural example-based adaptive tutoring systems. *IEEE Transactions on Education*, *46*(2), 241–251.

del Solato, T., & Du Boulay, B. (1995). Implementation of motivational tactics in tutoring systems. *Journal of Interactive Learning Research*, *6*(4), 337.

Dennis, M., Masthoff, J., & Mellish, C. (2012). The quest for validated personality trait stories. In *Proceedings of iui 2012* (pp. 273–276). ACM.

Dennis, M., Masthoff, J., & Mellish, C. (2016). Adapting progress feedback and emotional support to learner personality. *International Journal of Artificial Intelligence in Education*, *26*(3), 877–931.

Devin, H. F., Ghahramanlou, F., Fooladian, A., & Zohoorian, Z. (2012). The relationship between locus of control (internal–external) and happiness in pre-elementary teachers in iran. *Procedia-Social and Behavioral Sciences*, *46*, 4169–4173.

Dreher, C., Reiners, T., Dreher, N., & Dreher, H. (2009). Virtual worlds as a context suited for information systems education: Discussion of pedagogical experience and curriculum design with reference to second life. *Journal of Information Systems Education*, *20*(2), 211.

Escudero, E. B., Reyna, N. L., & Morales, M. R. (2000). The level of difficulty and discrimi-

nation power of the basic knowledge and skills examination (exhcoba). *Revista Electrónica de Investigación Educativa*, *2*(1), 2.

Fazlollahtabar, H., & Mahdavi, I. (2009). User/tutor optimal learning path in e-learning using comprehensive neuro-fuzzy approach. *Educational Research Review*, *4*(2), 142–155.

Forbes-Riley, K., Rotaru, M., & Litman, D. J. (2008). The relative impact of student affect on performance models in a spoken dialogue tutoring system. *User modeling and user-adapted interaction*, *18*(1), 11–43.

Gagne, R. M. (1968). Presidential address of division 15 learning hierarchies. *Educational psychologist*, *6*(1), 1–9.

Goldberg, L. (1980). Some ruminations about the structure of individual differences: Developing a common lexicon for the major characteristics of human personality. In *Invited paper, convention of the western psychological association, honolulu, hawaii*.

Graesser, A. C., Jeon, M., & Dufty, D. (2008). Agent technologies designed to facilitate interactive knowledge construction. *Discourse processes*, *45*(4-5), 298–322.

Graf, S., Lin, T., et al. (2008). The relationship between learning styles and cognitive traits–getting additional information for improving student modelling. *Computers in Human Behavior*, *24*(2), 122–137.

Grivokostopoulou, F., Perikos, I., & Hatzilygeroudis, I. (2015). Estimating the difficulty of exercises on search algorithms using a neuro-fuzzy approach. In *Tools with artificial intelligence (ictai), 2015 ieee 27th international conference on* (pp. 866–872).

Grivokostopoulou, F., Perikos, I., & Hatzilygeroudis, I. (2017). Difficulty estimation of exercises on tree-based search algorithms using neuro-fuzzy and neuro-symbolic approaches. In *Advances in combining intelligent methods* (pp. 75–91). Springer.

Heatherton, T. F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social psychology*, *60*(6), 895.

Hoffman, K. D., Hoffman, K. D., Kelley, S. W., Kelley, S. W., Rotalsky, H. M., & Rotalsky, H. M. (2016). Retrospective: tracking service failures and employee recovery efforts. *Journal of Services Marketing*, *30*(1), 7–10.

Holloway, R. L. (1978). Task selection and locus of control in two ability groups' recall. *Contemporary Educational Psychology*, *3*(2), 118–126.

Hospers, G.-J. (2003). Creative cities: Breeding places in the knowledge economy. *Knowledge, Technology & Policy*, *16*(3), 143–162.

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2002). Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct? *Journal of personality and social psychology*, *83*(3), 693.

Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel psychology*, *52*(3), 621–652.

Jugo, I., Kovacic, B., & Slavuj, V. (2014). Using data mining for learning path recommendation and visualization in an intelligent tutoring system. In *Information and communication technology, electronics and microelectronics (mipro), 2014 37th international convention on* (pp. 924–928).

Kelly, D. (2008). Adaptive versus learner control in a multiple intelligence learning environment. *J. of Educ. Multimedia and Hypermedia*, *17*(3), 307.

Kicken, W., Brand-Gruwel, S., & van Merriënboer, J. J. (2008). Scaffolding advice on task selection: a safe path toward self-directed learning in on-demand education. *Journal of Vocational Education and Training*, *60*(3), 223–239.

Kilpatrick, J. (2014). History of research in mathematics education. In *Encyclopedia of mathematics education* (pp. 267–272). Springer.

Koedinger, K. R., & Anderson, J. R. (1993). Effective use of intelligent software in high school math classrooms.

Kostons, D., van Gog, T., & Paas, F. (2010). Self-assessment and task selection in learner-controlled instruction: Differences between effective and ineffective learners. *Computers & Education*, *54*(4), 932–940.

Koutsojannis, C., Beligiannis, G., Hatzilygeroudis, I., Papavlasopoulos, C., & Prentzas, J. (2007). Using a hybrid ai approach for exercise difficulty level adaptation. *International Journal of Continuing Engineering Education and Life Long Learning*, *17*(4-5), 256–272.

Krause, J. S., & Broderick, L. (2006). Relationship of personality and locus of control with employment outcomes among participants with spinal cord injury. *Rehabilitation counseling bulletin*, *49*(2), 111–114.

Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of educational research*, *75*(1), 63–82.

Latham, A., Crockett, K., McLean, D., & Edmonds, B. (2012a). Adaptive tutoring in an intelligent conversational agent system. In *Transactions on computational collective intelligence viii* (pp. 148–167). Springer.

Latham, A., Crockett, K., McLean, D., & Edmonds, B. (2012b). A conversational intelligent tutoring system to automatically predict learning styles. *Computers & Education*, *59*(1), 95–109.

Lin, C., Liu, D., Pang, W., & Wang, Z. (2015). Sherlock: A semi-automatic framework for quiz generation using a hybrid semantic similarity measure. *Cognitive computation*, *7*(6), 667–679.

Llewellyn, D. J., Lang, I. A., Langa, K. M., & Huppert, F. A. (2008). Cognitive function and psychological well-being: findings from a population-based cohort. *Age and Ageing*, *37*(6), 685–689.

Lo, J.-J., Chan, Y.-C., & Yeh, S.-W. (2012). Designing an adaptive web-based learning system based on students cognitive styles identified online. *Computers & Education*, *58*(1), 209–222.

Lusk, D. L., Evans, A. D., Jeffrey, T. R., Palmer, K. R., Wikstrom, C. S., & Doolittle, P. E. (2009). Multimedia learning and individual differences: Mediating the effects of working memory capacity with segmentation. *British Journal of Educational Technology*, *40*(4), 636–651.

Magoulas, G. D., Papanikolaou, Y., & Grigoriadou, M. (2003). Adaptive web-based learning: accommodating individual differences through system's adaptation. *British journal of educational technology*, *34*(4), 511–527.

Malkoç, A. (2011). Big five personality traits and coping styles predict subjective well-being: A study with a turkish sample. *Procedia-Social and Behavioral Sciences*, *12*, 577–581.

Mampadi, F., Chen, S. Y., Ghinea, G., & Chen, M.-P. (2011). Design of adaptive hypermedia learning systems: A cognitive style approach. *Computers & Education*, *56*(4), 1003–1011.

Martinez, M., & Bunderson, C. V. (2000). Foundations for personalized web learning environments. *ALN Magazine*, *4*(2).

Maslow, A. H. (1973). *On dominance, self-esteem, and self-actualization*. Maurice Bassett.

Masthoff, J. (2006). The user as wizard: A method for early involvement in the design and evaluation of adaptive systems. In *5th workshop on user-centred design and adaptive systems* (pp. 460–469).

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of personality*, *60*(2), 175–215.

Mcquiggan, S. W., Mott, B. W., & Lester, J. C. (2008). Modeling self-efficacy in intelligent tutoring systems: An inductive approach. *User modeling and user-adapted interaction*, *18*(1-2), 81–123.

Meeks, B. S., Hendrick, S. S., & Hendrick, C. (1998). Communication, love and relationship satisfaction. *Journal of Social and Personal Relationships*, *15*(6), 755–773.

Melis, E., & Andres, E. (2005). Global feedback in activemath. *The Journal of Computers in Mathematics and Science Teaching*, *24*(2), 197.

Mitrović, A., Djordjević-Kajan, S., & Stomenov, L. (1996). Instruct: Modeling students by asking questions. *User Modeling and User-Adapted Interaction*, *6*(4), 273–302.

Mitrovic, A., Martin, B., & Mayo, M. (2002). Using evaluation to shape its design: Results and experiences with sql-tutor. *User Modeling and User-Adapted Interaction*, *12*(2), 243–279.

Miwa, K., Tanaka, M., Okazaki, S., Furukado, S., Yagita, Y., Sakaguchi, M., ... Kitagawa, K. (2014). Chronic kidney disease is associated with dementia independent of cerebral small-vessel disease. *Neurology*, *82*(12), 1051–1057.

Montazemi, A. R., & Wang, F. (1995). On the effectiveness of a neural network for adaptive external pacing. *J. of Interactive Learning Research*, *6*(4), 379.

Mora, V. A. Q., Gallego-Durán, F. J., Molina-Carmona, R., & Llorens-Largo, F. (2017). Subliminal learning. what do games teach us? In *International conference on learning and collaboration technologies* (pp. 487–501).

Moutafi, J., Furnham, A., & Paltiel, L. (2005). Can personality factors predict intelligence? *Personality and Individual Differences*, *38*(5), 1021–1033.

MT. (2012). *Amazon mechanical turk.* http://www.mturk.com.

Nicholls, J. G., & Miller, A. T. (1983). The differentiation of the concepts of difficulty and ability. *Child development*, 951–959.

Okpo, J. (2016). Adaptive exercise selection for an intelligent tutoring system. In *Proceedings of the 2016 conference on user modeling adaptation and personalization* (pp. 313–316).

Okpo, J., Dennis, M., Masthoff, J., Smith, K. A., & Beacham, N. (2016). Exploring requirements for an adaptive exercise selection system. In *Proceedings of the 6th workshop on personalization approaches in learning environments (pale 2016).*

Okpo, J., Dennis, M., Smith, K., Masthoff, J., & Beacham, N. (2016). Adapting exercise selection to learner self-esteem and performance. In *Intelligent tutoring systems* (pp. 517–518).

Okpo, J., Masthoff, J., Dennis, M., & Beacham, N. (2017). Conceptualizing a framework for adaptive exercise selection with personality as a major learner characteristic. In *Adjunct publication of the 25th conference on user modeling, adaptation and personalization* (pp. 293–298).

Okpo, J., Masthoff, J., Dennis, M., Beacham, N., & Ciocarlan, A. (2017). Investigating the impact of personality and cognitive efficiency on the selection of exercises for learners. In *Proceedings of the 25th conference on user modeling, adaptation and personalization* (pp. 140–147).

Oppermann, R., & Rasher, R. (1997). Adaptability and adaptivity in learning systems. *Knowledge transfer*, *2*, 173–179.

Paas, F. G., & Van Merriënboer, J. J. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *35*(4), 737–743.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of educational research*, *66*(4), 543–578.

Petrovica, S. (2013). Adaptation of tutoring to students' emotions in emotionally intelligent tutoring systems. In *e-learning and e-technologies in education (iceee), 2013 second international conference on* (pp. 131–136).

Pholo, D., & Ngwira, S. (2013). Integrating explicit problem-solving teaching into activemath, an intelligent tutoring system. In *Adaptive science and technology (icast), 2013 international conference on* (pp. 1–8).

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology*, *82*(1), 33.

Pon-Barry, H., Schultz, K., Bratt, E. O., Clark, B., & Peters, S. (2006). Responding to student uncertainty in spoken tutorial dialogue systems. *Int. J. of Artificial Intelligence in Education*, *16*(2), 171–194.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological bulletin*, *135*(2), 322.

Rammstedt, B., Danner, D., & Martin, S. (2016). The association between personality and cognitive ability: Going beyond simple effects. *Journal of Research in Personality*, *62*, 39–44.

Ravi, G. A., & Sosnovsky, S. (2013). Exercise difficulty calibration based on student log mining. *Proceedings of DAILE*, *13*.

Ray, R. D., & Belden, N. (2007). Teaching college level content and reading comprehension skills simultaneously via an artificially intelligent adaptive computerized instructional system. *The Psychological Record*, *57*(2), 201.

Read, J. P., Kahler, C. W., Strong, D. R., & Colder, C. R. (2006). Development and preliminary validation of the young adult alcohol consequences questionnaire. *Journal of studies on alcohol*, *67*(1), 169–177.

Reategui, E., Boff, E., & Campbell, J. A. (2008). Personalization in an interactive learning environment through a virtual character. *Computers & Education*, *51*(2), 530–544.

Reigeluth, C. M., & Carr-Chellman, A. (1999). Instructional-design theories and models (volume ii). *A new paradigmofinstructional theory*.

Rosenberg, M. (1986). *Conceiving the self*. RE Krieger.

Rosenberg, M., Schooler, C., Schoenbach, C., & Rosenberg, F. (1995). Global self-esteem and specific self-esteem: Different concepts, different outcomes. *ASR*, 141–156.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and applied*, *80*(1), 1.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161–1178.

Salden, R. J., Paas, F., & van Merriënboer, J. J. (2006). A comparison of approaches to learning task selection in the training of complex cognitive skills. *Computers in Human Behavior*, *22*(3), 321–333.

Salden, R. J., Paas, F., & Van Merriënboer, J. J. (2006). Personalised adaptive task selection in air traffic control: Effects on training efficiency and transfer. *Learning and Instruction*, *16*(4), 350–362.

Schwonke, R., Hauser, S., Nückles, M., & Renkl, A. (2006). Enhancing computer-supported writing of learning protocols by adaptive prompts. *Computers in Human Behavior*, *22*(1), 77–92.

Shute, V. J. (1995). Smart: Student modeling approach for responsive tutoring. *User Modeling and User-Adapted Interaction*, *5*(1), 1–44.

Soldz, S., & Vaillant, G. E. (1999). The big five personality traits and the life course: A 45-year longitudinal study. *Journal of Research in Personality*, *33*(2), 208–232.

Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation*, *2*(1), 50–80.

Stephens, A. C., Knuth, E. J., Blanton, M. L., Isler, I., Gardiner, A. M., & Marum, T. (2013). Equation structure and the meaning of the equal sign: The impact of task selection in eliciting elementary students understandings. *JMB*, *32*(2), 173–182.

Sun, P.-C., & Cheng, H. K. (2007). The design of instructional multimedia in e-learning: A media richness theory-based approach. *Computers & education*, *49*(3), 662–676.

Sun, P.-C., Tsai, R. J., Finger, G., Chen, Y.-Y., & Yeh, D. (2008). What drives a successful e-learning? an empirical investigation of the critical factors influencing learner satisfaction. *Computers & education*, *50*(4), 1183–1202.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Measuring cognitive load. In *Cognitive load theory* (pp. 71–85). Springer.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, *10*(3), 251–296.

Taminiau, E. M., Kester, L., Corbalan, G., Alessi, S. M., Moxnes, E., Gijselaers, W. H., … Van MerriëNboer, J. J. (2013). Why advice on task selection may hamper learning in on-demand education. *Computers in Human Behavior*, *29*(1), 145–154.

Taminiau, E. M., Kester, L., Corbalan, G., Spector, J. M., Kirschner, P. A., & Van Merriënboer, J. (2015). Designing on-demand education for simultaneous development of domain-specific and self-directed learning skills. *JCAL*, *31*(5), 405–421.

Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, *30*, 415433.

Tennyson, R. D. (1993). Mais: A computer-based integrated instructional system. *Behavior*

Research Methods, *25*(2), 93–100.

Tennyson, R. D., Thurlow, R., & Breuer, K. (1987). Problem-oriented simulations to develop and improve higher-order thinking strategies. *Computers in Human Behavior*, *3*(3), 151–165.

Triantafillou, E., Pomportsis, A., Demetriadis, S., & Georgiadou, E. (2004). The value of adaptivity based on cognitive style: an empirical study. *British Journal of Educational Technology*, *35*(1), 95–106.

Trudeau, F., & Shephard, R. J. (2008). Physical education, school physical activity, school sports and academic performance. *International Journal of Behavioral Nutrition and Physical Activity*, *5*(1), 1.

Tseng, J. C., Chu, H.-C., Hwang, G.-J., & Tsai, C.-C. (2008). Development of an adaptive learning system with two sources of personalization information. *Computers & Education*, *51*(2), 776–786.

Tsiriga, V., & Virvou, M. (2004). Evaluating the intelligent features of a web-based intelligent computer assisted language learning system. *International journal on artificial intelligence tools*, *13*(02), 411–425.

Vandewaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior*, *27*(1), 118–130.

Van Gog, T., Ericsson, K. A., Rikers, R. M., & Paas, F. (2005). Instructional design for advanced learners: Establishing connections between the theoretical frameworks of cognitive load and deliberate practice. *Educational Technology Research and Development*, *53*(3), 73–81.

Van Gog, T., Kester, L., & Paas, F. (2011). Effects of concurrent monitoring on cognitive load and performance as a function of task complexity. *Applied Cognitive Psychology*, *25*(4), 584–587.

Varela, F. J., Thompson, E., & Rosch, E. (2017). *The embodied mind: Cognitive science and human experience.* MIT press.

Verdú, E., Regueras, L. M., Verdú, M. J., Leal, J. P., de Castro, J. P., & Queirós, R. (2012). A distributed system for learning programming on-line. *Computers & Education*, *58*(1), 1–10.

Vinciarelli, A., & Mohammadi, G. (2014). A survey of personality computing. *IEEE Transactions on Affective Computing*, *5*(3), 273–291.

Watson, D., A, C. L., & Tellegen, A. (1998). Development and validation of brief measures of positive and negative affect: the panas scales. *J. of personality and social psychology*, *54*, 1063-1070.

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *J. of personality and social psych.*, *76*(5), 820.

Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement*, *39*(1), 1–37.

Wu, G., & Cheng, I. (2007). An interactive 3d environment for computer based education. In *Multimedia and expo, 2007 ieee international conference on* (pp. 1834–1837).

Xie, B., & Salvendy, G. (2000). Review and reappraisal of modelling and predicting mental workload in single-and multi-task environments. *Work & stress*, *14*(1), 74–99.