Metacognitive Aspects of Learning
What Influences Magnitude and Accuracy of Ease-of-Learning Judgments?

Andreas Jemstedt

Academic dissertation for the Degree of Doctor of Philosophy in Psychology at Stockholm University to be publicly defended on Friday 7 September 2018 at 13.00 in David Magnussonsalen (U31), Frescati Hagväg 8.

Abstract
To learn efficiently, many situations require people to judge what will be easy or difficult to learn, or how well it has been stored in memory. These metacognitive judgments are important to understand because they most likely guide how people behave when they learn, and consequently how much they learn. In this thesis, I focus on what is referred to as ease-of-learning (EOL) judgments, that is judgments about how easy or difficult a material will be to learn. EOL judgments have received relatively limited attention in the metacognitive literature. Therefore, this thesis also considers for comparison the more extensively researched judgments of learning (JOL), which are judgments of how well a studied material has been learned or how likely it is to be remembered on a later memory test. I had two major aims with my research. First, I aimed to investigate how accurate EOL judgments are, that is, how well they can predict the ease of future learning, and what moderates this accuracy. More precisely, I investigated what affects EOL judgment accuracy by varying how much an item-set varies in a predictive item characteristic, as well as varying methodological aspects of the judgment situation. The second major aim was to investigate what sources of information people use to make EOL judgments and how the information is used to make metacognitive judgments. In three studies, participants made EOL judgments for word pairs (e.g., sun – warm), or single words (e.g., bucket), studied the items, and tried to recall them on memory tests. In Study II, participants also made JOLs after studying the items. To estimate the accuracy of the judgments, the judgments were correlated with recall performance on memory tests. The results of the thesis show that EOL judgments can be accurate when they are made on a to-be-learned material which varies in a predictive item characteristic (Study I and II). In some conditions, EOL judgments are even as accurate as JOLs (Study II). Study II also supports the cue competition hypothesis, which predicts that, when people judge memory and learning, they sometimes rely less on one source of information if other information is available. Furthermore, Study III shows that processing fluency (the experience of effort associated with processing information), may be an important source of information for EOL judgments, and that people’s beliefs about available information can moderate how the information is used to make EOL judgments. Overall, the results show when EOL judgments will be accurate and when they will not, and provides evidence that people may use processing fluency to make EOL judgments even when it contradicts their beliefs. Importantly, the results also indicate that when multiple sources of information are available, information may compete for influence over metacognitive judgments.

Keywords: ease-of-learning judgments, judgments of learning, metacognition, monitoring, processing fluency, learning, memory.

Stockholm 2018
http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-157395

ISBN 978-91-7797-233-4

Department of Psychology
Stockholm University, 106 91 Stockholm
METACOGNITIVE ASPECTS OF LEARNING
Andreas Jemstedt
Metacognitive Aspects of Learning
What Influences Magnitude and Accuracy of Ease-of-Learning Judgments?

Andreas Jemstedt
För att kunna studera effektivt behöver människor kunna bedöma hur lätt eller svårt något kommer vara att lära sig, samt kunna bedöma hur väl de har lärt sig något. Om en student exempelvis missbedömer hur svåra olika kapitel i en kursbok kommer vara att lära sig kan det leda till att studenten planerar sin studietid på ett ineffektivt sätt. På samma sätt kan studenten studera olika kapitel av boken för länge eller för kort tid, om denne missbedömer hur väl denne lärt sig innehållet i de olika kapitlen. Inkorrepta bedömningar kan således sannolikt leda till att människor lär sig mindre, eller mindre effektivt. Bedömningar som dessa kallas metakognitiva bedömningar eftersom det är bedömningar av den egna kognitionen, det vill säga bedömningar av egna mentala processer. I min forskning har jag fokuserat på så kallade ease-of-learning (EOL) bedömningar, vilka är bedömningar av just hur lätt eller svårt något kommer vara att lära sig. Inom den metakognitiva litteraturen har EOL-bedömningar fått förhållandevis lite uppmärksamhet. På grund av detta har jag även jämfört dem med de mer omskrivna judgments of learning (JOL), vilka är bedömningar av hur väl något är lagrat i minnet.


Resultaten från studierna visar att EOL-bedömningar kan vara korrekta, men endast när det bedömda materialet varierar i egenskaper som låter bedömarendis särskilja mellan svåra och lätta ordpar (Studie I och II). I vissa fall var EOL-bedömningarna till och med lika korrekta som JOL-bedömningar (Studie II), trots att tidigare studier har antytt att JOL-bedömningar är mer korreka. Studie II antyder även att människor ibland förlitar sig mindre på en typ av information om annan information finns tillgänglig. Vidare visade Studie III att människor bedömer ord presenterade i en betingelse förknippad med
högre kognitiv ansträngning som svårare att lära sig, jämfört med när de presenterades i en betingelse förknippad med lägre kognitiv ansträngning. Hur stor skillnad det var mellan betingelserna modererades dock av hur deltagarna trodde att betingelserna skulle påverka deras förmåga att lära sig orden. Detta antyder alltså att den kognitiva ansträngningen tycks ha en direkt effekt på bedömningarna, eftersom deltagarnas EOL-bedömningar i många fall påverkades av ansträngningen i motsatt riktning mot vad de själva trodde.

Sammanfattningsvis visar avhandlingens tre studier att människor korrekt kan bedöma hur lätt eller svårt något kommer vara att lära sig. Bedömningarna kan dock vara inkorrepta om människor saknar information som de kan använda sig av, eller om informationen de använder sig av inte är prediktiv. Studierna belyser även hur människor använder sig av information för att göra metakognitiva bedömningar, genom att visa att människor ibland förlitar sig mindre på en informationskälla när en annan källa finns tillgänglig. En av studierna visar också att hur människor tror att en informationskälla påverkar deras inlärning, inte helt kan förklara hur källan påverkar deras bedömningar. Resultaten från studierna kan även vara användbara för att förstå lärande utanför labbet, eftersom de antyder att människor i många fall inte borde förlita sig på sina egna bedömningar när de planerar sin studietid.
List of studies

This doctoral thesis is based on the following three studies:


Contents

Sammanfattning på svenska ............................................................................ i
List of studies .................................................................................................. iii
Abbreviations .................................................................................................. vi
Introduction ..................................................................................................... 1
  What is metacognition? .................................................................................. 3
  How are metacognitive judgments investigated? .......................................... 6
    Ease-of-Learning Judgments ..................................................................... 8
    Judgments of Learning .......................................................................... 8
    Differences between metacognitive judgments ....................................... 9
    Measuring judgment accuracy ................................................................ 10
  What affects the magnitude and accuracy of EOL judgments? .................. 12
    Ease of learning judgments ..................................................................... 12
    How similar are EOL judgments to JOLs? .............................................. 15
  How do people make metacognitive judgments? ........................................ 16
    The cue-utilization approach .................................................................. 17
    How do people use multiple cues? .......................................................... 18
    Processing fluency .................................................................................. 20
    Metacognitive beliefs ............................................................................. 21
  The aims of this thesis ............................................................................... 23

Empirical studies .......................................................................................... 24
  Study I: Jemstedt, Kubik, and Jönsson (2017) .............................................. 24
    Experiment 1 .......................................................................................... 24
    Experiment 2 .......................................................................................... 28
    Experiment 3 .......................................................................................... 28
  Study II: Jönsson, Jemstedt, Schwartz, and Kubik (2017) ......................... 30
    Experiment 1 .......................................................................................... 31
    Experiment 2 .......................................................................................... 34
  Study III: Jemstedt, Schwartz, and Jönsson (2017) .................................... 36
    Experiment 1 and 2 ................................................................................ 36

General Discussion ....................................................................................... 41
  The accuracy of EOL judgments .............................................................. 41
  How does the discriminability of a predictive cue affect EOL judgment accuracy? 42
Abbreviations

- EOL judgments – Ease of learning judgments
- FOK judgments – Feeling of knowing judgments
- JOL – Judgment of learning
Introduction

Learning is an essential part of human life. In almost every situation in life, knowledge from prior experiences affects how people behave. Although skills, behaviors and knowledge are often learned without people actively trying to learn them, many things are learned because people had the goal to learn. Active learning may include anything from studying for a history exam, learning a new language, or practicing playing the piano. Although many aspects of learning impact how efficiently people learn, an important aspect is knowing how to achieve a goal, or knowing how easy a to-be-learned material will be to learn. That is, for efficient learning, a person benefits from knowing how and what to study, and for how long. This aspect of learning involves what is called metacognition. In short, metacognition can be described as thinking about thinking, and more formally it has been described as monitoring and control of one’s own cognition (Nelson & Narens, 1990). This aspect of learning is the focus of this doctoral thesis. Specifically, it focuses on how people judge what will be easy or difficult to learn and how accurate these judgments can be.

The importance of understanding people’s thoughts about their own learning and memory was demonstrated in one of the earliest experiments on metacognition. Flavell, Friedrichs, and Hoyt (1970) asked children from nursery school up to fourth grade to study a set of pictures with familiar objects until all of the objects were memorized. After studying the pictures, the children were asked to recall the names of all objects in the pictures. When older children said that they had learned all the objects, this was indeed the case for most of them. However, when younger children said they were ready, most of them failed to recall all of the objects. This shows the importance of knowing when something has been learned, as one may otherwise stop studying prematurely. Importantly, later laboratory studies showed that although adults can make accurate judgments about their current state of learning, they, like the children, might erroneously think that they learned something that they had not (e.g., Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991; for a review and meta-analysis see, Rhodes, 2016; Rhodes & Tauber, 2011). More importantly for this thesis, adults can also mistakenly judge difficult materials as being easy to learn, or vice versa (e.g., Leonesio & Nelson, 1990).
Although this doctoral thesis focuses on metacognition in adults, the results from Flavell et al. (1970) is a good example of the consequences which inaccurate metacognition can have for learning. These insights from research indicate that, in everyday learning situations, people may sometimes be unaware of, or misjudge, what will be easy or difficult to learn and whether they have reached their learning goals or not. In a worst-case scenario, failing to accurately judge what one has learned, or how difficult a material will be to learn, could result in students failing a test, or studying a chapter in a course book even though all of the information is already stored in memory, thus wasting valuable time. Therefore, investigating metacognition to understand how people judge their learning or memory and how these judgments affect behaviors is an important aim for researchers. Such research may be valuable for improving learning, both when developing efficient learning strategies that can be taught in schools and also for people practicing alone. Although metacognition plays an important role in learning (Nelson & Narens, 1990), it is still unclear exactly how people monitor their cognitive processes.

In my research, my primary focus is on ease-of-learning (EOL) judgments. These are judgments of what will be easy or difficult to learn. To date, the literature on EOL judgments has been fairly limited, as also noted by Dunlosky and Metcalfe (2009, p. 91), who wrote that “sustained devotion to understanding EOL judgments has been rather minimal and little is known about them.” This description of the field still holds true, as little research has focused on EOL judgments in the last decade. Hopefully, the research presented in this doctoral thesis will increase the understanding of how EOL judgments work, as well as contribute to the growing field of metacognition.

The thesis had two major aims with related research questions. The first major aim was to investigate how accurate EOL judgments are and what moderates the accuracy. The second aim was to investigate what information people use to make EOL judgments, and how people use information to make metacognitive judgments.

In the following sections of this introduction, I will first introduce the reader to what metacognition is and how it is commonly investigated. Following this, I will focus on EOL judgments and what determines if people judge an item as being easy or difficult to learn, and what affects the accuracy of these judgments. I will then review the existing literature on EOL judgments and describe possible similarities and differences between EOL judgments and the more widely investigated judgments of learning (JOL), which are judgments about how well material has been learned after studying it, or of how likely one is to remember a studied material on an upcoming memory test. Thereafter, I will describe the current theories on how people make metacognitive judgments. Finally, I will describe the research questions my studies aimed to investigate in more detail.
What is metacognition?

One of the primary components of metacognition is subjective judgments of cognitive events (metacognitive monitoring), and these judgments most likely influence how people behave (metacognitive control; Nelson & Narens, 1990). In one influential paper on metacognition, Flavell (1979) argued for a model of metacognitive monitoring and related phenomena. He argued that metacognitive monitoring takes place by means of four classes of phenomena and the interactions between them, namely, (I) metacognitive knowledge, (II) metacognitive experiences, (III) goals and tasks, and (IV) actions and strategies. Specifically, he suggested that, when people monitor their memory and learning, (I) they do so by applying the beliefs and knowledge they have about their own cognition. For example, people may believe that they are good at taking tests, learning languages, or that long words are more difficult to learn than short ones (cf., Mueller & Dunlosky, 2017). People may also believe that one study strategy is more efficient than another. These beliefs then influence how they judge their learning, and consequently influence their behavior. At the same time, people may also monitor (II) experiences associated with how well an ongoing task is progressing, or perhaps the feeling that one will fail an upcoming test. For example, if a previously studied word is easily retrieved from memory, this may signal that the word is well stored in memory (e.g., Metcalfe & Finn, 2008b). Accordingly, a person may judge a word that is easily remembered as better memorized than a word that is more effortful to remember. Monitoring can also be affected by (III) the aimed-for outcomes of current cognitive efforts. If a person has been given the task of learning the content of a textbook chapter, or has the goal of memorizing a grocery list for later retrieval in a store, the goal or task may influence which aspects of the learning situation will influence monitoring. For example, if one item on a grocery list is more important to remember, people may spend more time memorizing that item (cf., Ariel, Dunlosky, & Bailey, 2009). Finally, monitoring can also be affected by (IV) which techniques or approaches people apply to reach their goals or complete their tasks. For example, if people test themselves during learning to strengthen their memory (Roediger & Karpicke, 2006), the outcome of the test also signals what is stored in memory (e.g., Dunlosky & Nelson, 1992), helping them to judge what they have learned or not. In my research, I have primarily focused on what Flavell (1979) called metacognitive knowledge (I) and metacognitive experiences (II).

Flavell’s (1979) model of cognitive monitoring is still an important basis for understanding and investigating metacognition. Another influential theoretical contribution to the study of metacognition is the theoretical framework by Nelson and Narens (1990). In their framework, they describe how cognitive processes are divided into two separate, but connected, levels (See Figure 1) – an object-level and a meta-level. The object-level is the currently ongoing processes such as memory retrieval or learning. At the same time, the meta-
level contains a dynamic, but often simplified, model of the object-level processes. The two levels communicate via processes called monitoring and control. When information flows from the object-level to the meta-level, it is called monitoring. This information is used to update and change the dynamic model. Based on the current dynamic model, information from the meta-level controls the object-level. Control either makes changes to the process that the object-level is currently working on, orders it to change from one process to another, or simply orders it to continue what it is doing. For example, if a student is trying to learn Ukrainian by studying a list of Swedish-Ukrainian word pairs, the object level works on memorizing the pairs. Meanwhile, the meta-level monitors the object-level using available information to create a model of how the learning is progressing. Based on the meta-level model, the object-level can be controlled to switch from studying one pair to another, if a pair is judged as learned, or to stay on the current pair if not.

Figure 1. The figure illustrates how information flows from an object-level to a meta-level (monitoring), and how information flows from the meta-level to an object-level (control). Adapted from Metamemory: A Theoretical Framework and New Findings (p. 126), by Nelson, T. O., and Narens, L., 1990, The psychology of learning and motivation (Vol. 26), New York, NY: Academic Press. Copyright 1990 by Academic Press Inc. Adapted with permission.

An important aspect of this framework is that monitoring often fails to perfectly convey the status of the ongoing object-level process. This imperfect monitoring may occur because information is misinterpreted, or the information is of poor quality. In the earlier example, this could result in the student
judging that a word pair is memorized when it is not, or vice versa. At the same time, because control is based on the current model of the object-level, a faulty model can result in sub-optimal control. It should also be noted that, although the control process does not directly give the meta-level any information about the object-level, experiments have shown that, under some conditions, feedback flowing from the object-level can be used to update the model in the meta-level (Koriat, Ma’ayan, & Nussinson, 2006). For example, if the student studies one word pair longer than another pair, this information can be used to judge the pair as more difficult to remember than the pair that was studied for a shorter time. When people make EOL judgments, these judgments are most likely being based on the current meta-level model of how easy or difficult the current item may be to learn. That is, when the object-level processes an item, information about the item flows to the meta-level and is used to construct a model of how easy or difficult the item may be to learn. Based on this model, people then make an EOL judgment. The information used to construct the model could be, for example, how effortful a word is to process (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989) or how concrete a word is (Jönsson & Lindström, 2010).

How then do people utilize these different sources of information to make metacognitive judgments? It is well established that people judge their learning or memory by making inferences about their cognition based on how available information affects their learning and memory (e.g., Koriat, 1997). For example, people may use the length of a word to infer how easy or difficult it will be to learn (Jönsson & Lindström, 2010). More precisely, if words vary in length, people may infer that the longer words will be more difficult to learn than shorter words and base their EOL judgment on this. Henceforth, I will refer to the available information people use to make judgments as cues, because this is the term used in the metacognitive literature. Towards the end of this introduction, I will, in more detail, discuss how people may use cues to make inferences about their learning and memory.

This thesis focuses on metacognitive monitoring and the judgments based on the monitoring. One reason it is important to understand this aspect of metacognition is that metacognitive control most likely guides people’s study behavior (i.e., metacognitive control; Nelson & Narens, 1990). Importantly, during the last decade, experimental research has demonstrated that JOLs are causally related to study decisions made after studying (e.g., Metcalfe & Finn, 2008a; Rhodes & Castel, 2009; Soderstrom & Rhodes, 2014). These findings likely also generalize to monitoring at other times in the learning process (e.g., EOL judgments). To be precise, if people can accurately judge which items will be easy and difficult to learn, this should help them to study efficiently.
How are metacognitive judgments investigated?

In Nelson and Narens’ (1990) theoretical framework, they also described which metacognitive judgments can be used to investigate metacognitive monitoring at different stages in the learning process. Moreover, they described how monitoring is related to control in each stage. They organized the learning process into three stages: acquisition, retention and retrieval (see Figure 2). It should be noted that there are more types of metacognitive judgments than those described in Figure 2, but I will only describe those judgments, as they are the most relevant for this thesis. Furthermore, in their framework, they refer to metacognition related to memory as metamemory, and this term is frequently used in the literature, but, for simplicity, I will use the more general term of metacognition throughout this thesis.

**Figure 2.** Metacognitive monitoring and control along the learning phases of acquisition, retention, and retrieval. The figure illustrates monitoring and control divided into three stages and the metacognitive judgments used to measure monitoring at each stage (top), and how monitoring affect control (bottom). Adapted from Metamemory: A Theoretical Framework and New Findings (p. 129), by Nelson, T. O., and Narens, L., 1990, *The psychology of learning and motivation* (Vol. 26), New York, NY: Academic Press. Copyright 1990 by Academic Press Inc. Adapted with permission.
In order to investigate monitoring, researchers ask people to make metacognitive judgments – subjective judgments resulting from people’s introspections – and then use these judgments as estimates of people’s monitoring. That is, when people are asked to make judgments about their learning and memory, their judgments are most likely based on their metacognitive monitoring. As can be seen in Figure 2, different metacognitive judgments are made depending on which stage in the learning process researchers are interested in. In the beginning of learning, before being instructed to study, people may make EOL judgments to assess how easy or difficult the to-be-learned material will be to learn. These judgments likely affect what study strategies people decide to use, and how they plan their study time. During study or after studying the material, people may make JOLs. That is, they judge how well they have learned the material or how likely they are to remember the material on an upcoming memory test. These judgments likely influence whether people continue with or stop studying parts of the material, and whether or not they decide to change their study strategy. I will return to these two types of judgments below.

The other two judgments in Figure 2 – not under scrutiny in this thesis – are feeling-of-knowing (FOK) judgments and retrospective confidence judgments. Like JOLs, FOK judgments are used to measure metacognitive monitoring during or after learning. FOK judgments are commonly predictions made on unretrieved items about how likely they are to recognize the correct item in a list of alternatives. In many experiments, no active studying has occurred before FOK judgments are made. Instead, FOK judgments have been made on general information questions, for example – “what is the capital of Australia?” (e.g., Hart, 1965; Schwartz, Boduroglu, & Tekcan, 2016). However, FOK judgments are sometimes also made after studying word pairs (e.g., Metcalfe, Schwartz, & Joaquim, 1993). Like JOLs, FOK judgments likely affect whether or not people stop studying a material or not. Furthermore, during memory retrieval, FOK judgments likely affect when people try to search their memory for an answer and when they give up (Reder & Ritter, 1992; Koriat & Levy-Sadot, 2001).

After memory retrieval, people may also judge how confident they are that their answers are correct by making retrospective confidence judgments. During a memory test, retrospective confidence judgments likely guide when people stop trying to search their memory for an answer. It may also guide whether people are satisfied with their answer or decide to search for more information to answer a question.

In metacognitive research, judgments are typically made either on an item-by-item basis, with one judgment for each item in a material, or over a set of items by making global judgments (also called aggregate judgments). When global judgments are made, participants are asked to judge, for example, how many items they will recall from a material (e.g., Mazzoni & Nelson, 1995).
Because I have focused on item-by-item judgments, I refer to this type of judgment when I use the term judgments in this thesis.

Ease-of-Learning Judgments

EOL judgments, the primary focus of this thesis, are made at the beginning of learning, before any instructed study has begun. These EOL judgments are defined as judgments before learning an item (Nelson & Narens, 1990), and as being predictions about what will be easy or difficult to learn. However, this definition may be problematic. When people make EOL judgments, they make them with the item available, which means that some learning is likely to occur when the judgment is made. Therefore, I prefer a slightly altered definition of EOL judgments. Namely, I define EOL judgments as judgments made early in the learning processes, before active (i.e., instructed) study has begun, about what will be easy or difficult to learn. I prefer this definition, as it does not assume that a judgment is made before learning.

In a typical experiment investigating EOL judgments, participants are presented items one at a time, and, for each item, they are asked to make a judgment by answering a question, for example, “How likely is it that you will learn this item for the upcoming test”. To answer the question, participants typically select the likelihood on a provided scale (e.g., Kelemen, Frost, & Weaver, 2000). However, the experimental procedure has varied between studies. For instance, in some experiments, participants instead rank items from least to most likely to be learned, or vice versa (e.g., Leonesio & Nelson, 1990). Because few studies have investigated EOL judgments, it is difficult to say how different procedures affect the judgments. However, in Study I of this thesis, I compared the outcomes of three methodological differences to see how they affect the accuracy of EOL judgments.

Throughout my research, I asked participants to make EOL judgments on word pairs (e.g., sun – warm; Study I and II), and instructed them that, after the judgments are completed, they would study the pairs to learn the association between the words. They were also instructed that they would be tested after study by having to recall the second word when presented with the first. Although not perfect, this simulates learning a new language, when people need to learn the translation of a word. In addition, I used learning of single words (Study III), which simulates, for example, memorizing a grocery list. I used these materials because they are the most common materials in metacognitive research, and because they allow more control over which properties in the material that may vary, compared to other, more complex, materials.

Judgments of Learning

If researchers are instead interested in how well people monitor how likely an item will be remembered later or how well it is stored in memory, they can
ask participants to make JOLs. JOLs are defined as judgments about future memory performance, made during or after acquisition (Nelson & Narens, 1990; for a review see Rhodes, 2016). These judgments are similar to EOL judgments, but they are typically made after instructed study has occurred. Participants are asked to judge, for example, “How confident are you that you will remember this item on the upcoming test?”. Similar to EOL judgments, JOLs are typically made on a scale, but the methods vary for JOLs as well (e.g., Leonesio & Nelson, 1990). When metacognitive monitoring occurs during study it can influence when people stop studying, either because they judge the material is stored in memory or that it is too difficult to learn (Figure 2). In a later section, I will describe EOL judgments and JOLs and their similarities and differences in greater detail.

Differences between metacognitive judgments
As described by Nelson and Narens’ (1990) theoretical framework (see Figure 2), different judgments are used depending on what type of monitoring researchers are interested in. In the beginning of learning, people can predict future learning (i.e., EOL judgments), during or after study people can instead predict future memory performance (i.e., JOLs), or whether or not an answer is stored in memory, even when it cannot be accessed at the time (i.e., FOK judgments). Finally, after retrieval, people can judge the correctness of their responses (i.e., retrospective confidence judgments). One of the main reasons these judgments will differ, is what cues they rely on (see Schwartz, 1994). That is, if each of the judgments rely on the same cues in the same way, an item judged as being easy to learn would also be judged as likely to be remembered on a future test, or as being the correct answer after retrieval. However, this is not always the case. Instead, an item judged as being easy to learn can, after study, be judged to be unlikely to be remembered on a future test. Most likely, this is because metacognitive judgments throughout the learning process have access to and rely on different cues. In other words, at different times and depending on what is judged, some cues are not available, or some of the available cues are invalid because they are not predictive of what is judged. For example, the concreteness of a word may be a valid cue for EOL judgments, but it may be invalid for retrospective confidence judgments, although the cue is available at both times. Furthermore, after study, JOLs can rely on the outcome of a retrieval attempt as a cue, but this is not possible for EOL judgments, as items have not yet been learned. Consequently, what cues are available and how people use them, will also have an impact on how accurate the judgments are.
Measuring judgment accuracy

To measure how accurately metacognitive judgments predict future learning and memory performance, participants' judgments are compared with a criterion measure. The criterion measure is typically the binary outcome on a memory test (e.g., Dunlosky & Nelson, 1992); namely, whether each item was correctly recalled or not. Consequently, for judgments to be accurate they should reflect the outcome of the criterion measure. For EOL judgments, this means that, after judging and then studying a set of items, items which have been judged to be easy to learn should have a higher chance of being remembered on a later test than items which are judged as difficult to learn. It should also be noted that other criterion measures have also been used, for example, how many study and test trials were needed before an item is first recalled (e.g., Leonesio & Nelson, 1990), or if a participant correctly recognized an item on a recognition test (e.g., Koriat & Levy-Sadot, 2001).

The accuracy of metacognitive judgments likely depends on mainly three components: what cues are used to make the judgments, how predictive the cues are of the performance on the memory test, and how well people translate the cues into judgments. For example, imagine a student who has studied a list of word pairs for a test. When judging how likely each pair is to be learned, the student may use the semantic relatedness between the two words as a cue by judging pairs with low relatedness (e.g., baby – north) as less likely to be remembered than pairs with high relatedness (e.g., sun – warm). Because semantic relatedness is related to learning and memory (e.g., Dunlosky & Matvey, 2001), it will affect the performance on the memory test (i.e., the criterion measure), and accuracy will be high. However, if we imagine that semantic relatedness has no influence on the criterion measure, accuracy will be low. Furthermore, some people may be better at translating differences in semantic relatedness into judgments than others. For example, if a list of word pairs consists of pairs with minimal, low, moderate, and high semantic relatedness, one student may judge pairs as easier to learn for every level of semantic relatedness. That is, pairs with high semantic relatedness are judged as easier to learn than pairs with moderate, followed by low, and then minimal semantic relatedness. Meanwhile, a second student may give the same judgment to pairs with high and moderate semantic relatedness, and the same judgment to pairs with low and minimal semantic relatedness. In the example, the first student is more sensitive to differences in the cue compared to the second student. This would likely result in the first student making more accurate judgments than the second student. Although it is well established that metacognitive judgments are based on cues (e.g., Koriat, 1997), and judgment accuracy depends on how predictive the used cues are of the criterion measure (e.g., Dunlosky & Nelson, 1992), it is still unclear whether some people are better at translating cues than others (e.g., Kelemen et al., 2000).
Calibration and Resolution

In metacognitive research, there are two main categories of accuracy, which are most commonly referred to as calibration and resolution. These categories are also referred to as absolute and relative accuracy, respectively. Importantly, calibration and resolution describe different aspects of judgment accuracy and are independent of each other. In my research, I have focused on resolution, but I will here briefly describe the two types of accuracy and how they differ.

Calibration refers to the correspondence between the average judgment and average performance. Calibration measures are used to estimate if participants are overconfident, underconfident, or if they are perfectly calibrated in their judgments. To do this, the difference between judgments on a set of items is compared with how many of the same items were recalled or recognized on a memory test. For example, if a participant is perfectly calibrated, only 40% of all items that are judged with a likelihood of 40% to be remembered will be remembered on the test. However, if 60% of the items are remembered on the test, the participant is underconfident, and if only 20% are remembered, the participant is overconfident. Thus, calibration estimates the difference between judgments and performance on the criterion measure.

Whereas calibration captures the match between judgments and criterion measure on a set of items, resolution measures if more items judged with a higher likelihood are remembered than items judged with a lower likelihood. Thus, resolution measures people’s ability to discriminate between items.

Several measures have been used to measure resolution, but typically it is measured with a correlation. Although several different measures of resolution have been used, the most common measure in metacognitive research is the non-parametric Goodman-Kruskal gamma correlation. The gamma correlation ranges from -1 to +1. If items with high judgments are recalled and items with low judgments are not recalled, the gamma correlation will be high (or the reverse, depending on how the judgment question is phrased). If instead the items that were not recalled received higher judgments, the correlation will be low. When the correlation is close to 0, it indicates that there is no relation between judgments and subsequent performance on the memory test. The reason for why the gamma correlation is widely used is because Nelson (1984) compared several types of accuracy measures and argued that the gamma correlation was the best available one. His main arguments for using the gamma correlation are that it makes no assumptions about scaling above ordinal level and that judgments and criterion measures can have more than two levels. However, the gamma correlation has been criticized for being biased, and alternative measures have been proposed (e.g., Benjamin & Diaz, 2008; Masson & Rotello, 2009). Nevertheless, because there is still no consensus on which alternative measure would be preferable to use, I have used gamma correlations as the measure of resolution in my research. In addition, although there
may be some problems with gamma correlations, I wanted to use the most common measure to compare my findings with prior research.

It is also important to note calibration and resolution are independent. A person can have perfect resolution, but at the same time be highly underconfident, or vice versa. By this, I mean that a person could correctly judge some items as being easier (e.g., judged to have a 90% chance of being learned) and other items as less likely to be learned (e.g., 70% chance), resulting in a high resolution. But at the same time the average judgment (e.g., 80%) could be much higher than average recall (e.g., 30%), indicating overconfidence. Similarly, judgments could be negatively correlated with recall but at the same time calibration could be perfect. Hence, depending on what researchers are interested in, they use either calibration or resolution or both in their research. Because I have focused on resolution in my research, from here on, when I use the term accuracy, I refer to resolution.

What affects the magnitude and accuracy of EOL judgments?

The two major aims for my research are related to what cues EOL judgments rely on and what affect the accuracy of the judgments. Although a great deal of cues have been identified to affect EOL judgments, it is still unclear what cues people rely on. Furthermore, studies measuring EOL judgment accuracy have reported varying accuracy, and it has not been systematically investigated why the accuracy varies between studies. It is therefore difficult to conclude how accurate EOL judgments can be, and whether they can be as accurate as other metacognitive judgments. To investigate these issues, across the three studies included in this thesis, I explore both potential cues which people may be using to make EOL judgments, and methodological aspects that could affect their accuracy. In this section, I describe what previous research has found to influence the magnitude and accuracy of EOL judgments and how similar or different EOL judgments may be to the much more extensively researched JOLs.

Ease of learning judgments

Because EOL judgments are made early in the learning process, they should be more reliant on characteristics of the to-be-learned material than judgments made during or after instructed study. Meanwhile, judgments made after study can also rely on cues signaling how well the material has been stored in memory (e.g., Koriat, 1997). Characteristics in the material can, for example, be the length of a word, how frequently it occurs in a language, or how concrete or abstract the word is (Jönsson & Lindström, 2010). However, because
people are exposed to the item when they judge it, it is possible that they may also be monitoring how easily or fast they are encoding the items to memory (see, Hertzog, Dunlosky, Robinson, & Kidder, 2003).

Because of the central role of EOL judgments in the current thesis, I will, in this section, review the literature on the few studies that exist. Specifically, I will review how different types of cues affect the magnitude and accuracy (i.e., resolution) of EOL judgments. When I selected studies for the review, I used the following three inclusion criteria. First, EOL judgments were made by judging how easy or difficult items would be to learn or study, how fast items would be to learn, or how likely items were to be learned during study. Second, EOL judgments had to be made on an item-by-item basis. Studies that used global EOL judgments were therefore excluded (e.g., Carroll & Korukina, 1999). Third, the judged item had to be present during the judgment. The third criterion excluded studies on the otherwise fairly similar pre-study JOLs (e.g., Castel, 2008). Pre-study JOLs are judgments made immediately before studying an item. For example, participants are told that the upcoming item will be presented in a large or small font and judge the item on this information alone (Mueller, Dunlosky, Tauber, & Rhodes, 2014).

What affects the magnitude of EOL judgments?

To the best of my knowledge, the first study on EOL judgments was conducted by Richardson and Erlebacher (1958). They asked participants to judge how fast they would be able to learn pairs of adjectives (e.g., first – new), nonsense syllables (e.g., hod – bof) or consonant syllables (e.g., hnk – gbr). The pairs were constructed to have either high or low meaningfulness, based on the first part of the pair (e.g., hod) and on the association strength between the first and second part (e.g., hod – bof). The results showed that participants’ judgments were related to the meaningfulness of the pairs, in that pairs with higher meaningfulness were judged to be learned faster. Following Richardson and Erlebacher (1958), the investigation of what types of cues people use when they make EOL judgments has shown that many different cues may be used, depending on the to-be-learned material. Specifically, research has shown that EOL judgments increase with word frequency, how easy words are to imagine (Begg et al., 1989; Jönsson & Lindström, 2010), the semantic relatedness between words in word pairs (Dunlosky & Matvey, 2001), how typical an item is of a category (Mazzoni, Cornoldi, Tomat, & Vecchi, 1997), and decrease with word length (Jönsson & Lindström, 2010).

Furthermore, aspects other than characteristics in the material have also been shown to affect EOL judgments. In one study, word pairs with high semantic relatedness were judged as being easier to learn when they were judged after items with lower semantic relatedness, compared to when they were judged without first judging the items with lower relatedness (Dunlosky & Matvey, 2001). This order effect may indicate that EOL judgments could be affected by an anchoring effect (e.g., Wilson, Houston, Etling, & Brekke,
That is, people may use their initial judgments as anchors and adjust their later judgments in relation to the earlier ones. Similarly, de Carvalho Filho and Yuzawa (2001), showed that if people are informed about other people’s memory performance on the items they are judging, their EOL judgments are influenced in the same direction as the information they are given.

EOL judgment may also be affected by how a to-be-learned material is presented at the time the judgment is to be made. For music, samples of music are judged as easier to learn when they are both read and listened to, compared to only read, followed by only listened to (Peynircioğlu, Brandler, Hohman, & Knutson, 2014). However, whether this generalizes to other types of materials than music has never been tested. Furthermore, Löffler, von der Linden and Schneider (2016) investigated how age and expertise affected EOL judgments made on soccer-related word pairs. Experts were participants with knowledge about soccer. Although analyses of EOL judgment magnitude were not reported in the study, E. Löffler (personal communication, November 16, 2017) confirmed young children gave higher EOL judgments than older age groups, even though they did not recall more of the items. This indicates that perhaps young children are more confident in their judgments. She also confirmed experts gave higher EOL judgments than non-experts, but experts also recalled more of the items. This indicates that both age and how proficient a person feels on a subject impact their judgments on how easy a material will be to learn.

In summary, the literature of EOL judgments shows that people can use characteristics of the material, how the material is presented, and their expertise, to infer how easy or difficult a material will be to learn. In addition, EOL judgments may be influenced by anchoring effects.

What affects the accuracy of EOL judgments?

As mentioned above, people use a wide variety of cues when they make EOL judgments, but, for the judgments to be accurate, the cues that people employ need to be predictive of learning, and people need use them correctly.

A few studies have investigated what affects the accuracy of EOL judgments, and, when comparing the studies, it is apparent that accuracy varies a great deal. The main reason for this seems to be that studies use different material compositions. When comparing studies with different material compositions, it is clear that when a material has a low variability in a predictive characteristic, EOL judgments tend to be inaccurate (Jönsson & Kerimi, 2011, Experiment 2; Kelemen et al., 2000; Leonesio & Nelson, 1990; Mazzoni et al., 1997). Meanwhile, if the material varies in some predictive characteristic, EOL judgments can be moderately or highly accurate (Jönsson & Kerimi, 2011, Experiment 1; Lippman & Kintz, 1968; Peynircioğlu et al., 2014; Underwood, 1966). In other words, when a predictive item characteristic varies
between items, people are able to discriminate between easy and difficult items, but, when all items are similar in the characteristic, this is harder.

However, a predictive cue that does not vary could potentially still be used to increase calibration accuracy. That is, over a set of items people may still be more calibrated on a material that includes a predictive cue compared to one without a predictive cue, even though their resolution is inaccurate for both materials. For example, if a material only consists of word pairs with high semantic relatedness, the average judgment will likely be higher than if all pairs have low semantic relatedness (c.f., Dunlosky & Matvey, 2001). However, because the cue does not vary, people may not be able to discriminate between items resulting in a low resolution (e.g., Jönsson & Kerimi, 2011).

In summary, the accuracy of EOL judgments will most likely be high if the available cues are predictive of the criterion measure and people can use the cue to discriminate between items in the material, and lower if they cannot. However, this has never been directly tested, and it is only apparent when comparing EOL judgment accuracy between studies using different material compositions. This issue is important to investigate because if studies do not take the material composition into account, it may lead to incorrect conclusions about the accuracy of EOL judgments. Therefore, in line with my first major aim, this was investigated in Study I.

How similar are EOL judgments to JOLs?

Compared to EOL judgments, there is extensive literature on JOLs (for reviews see, Rhodes, 2016; Rhodes & Tauber, 2011). Considerable research has been directed at the factors or cues that contribute to JOL magnitude and accuracy. In fact, people rely on similar cues when they make JOLs as when they make EOL judgments. For example, studies show people rely on characteristics in the material such as semantic relatedness between the words in word pairs (e.g., Arbuckle & Cuddy, 1969; Koriat, 1997) and how concrete words are (e.g., Hertzog et al., 2003). However, because JOLs occur after study, the cue landscape has changed in comparison to when EOL judgments are made, and people also have access to cues not available in the beginning of learning (i.e., EOL judgments). During study, some items may be better learned than others, making additional idiosyncratic cues (i.e., cues that will differ between individuals) available. Therefore, people may also rely on a feeling of familiarity when they are presented with the items, and, in some conditions, people may also rely on how easily an item is retrieved from memory (e.g., Metcalfe & Finn, 2008b; Nelson & Dunlosky, 1991).

The accuracy of JOLs depends on the conditions when the JOL is made. More precisely, JOL accuracy is influenced by when the judgment is made and how much of the material is available when the judgment is made (Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991). It is well established that delayed JOLs (a delay between studying and judging items) are more accurate
than immediate JOLs (a judgment is made directly after studying each item), and especially if people can make a diagnostic retrieval attempt when they make the judgment (for a meta-analysis see, Rhodes & Tauber, 2011). For example, when people have studied word pairs (e.g., book – chair), and the JOLs are made after a delay, and only the first word in each pair is visible (e.g., book – ?), their judgements become highly accurate (Nelson & Dunlosky, 1991). Although several explanations for this high accuracy have been proposed (e.g., Nelson & Dunlosky, 1991; Sikström & Jönsson, 2005), each an explanation is tied to people being able to make a diagnostic retrieval attempt when they make the JOLs. In other words, whether or not a word can be retrieved from memory is predictive of later retrieval and, when people can rely on this cue, it makes their judgments highly accurate.

Overall, JOLs are similar to EOL judgments, as both judgments are predictions about future memory, although one judgment is made before study and the other after study, and the judgment question differs slightly. Although JOL and EOL judgments are similar, they may differ in accuracy because the cue landscape has changed, and JOLs have access to the additional idiosyncratic cues signaling how well the material has been learned during study. However, to what degree these additional cues result in differences in accuracy has received only limited attention. Nevertheless, when one study compared EOL judgments and JOLs made with both the cue and target words visible (e.g., book – chair), the authors concluded that JOLs were more accurate than EOL judgments (Leonesio & Nelson, 1990). However, the study used a homogeneous material in which all the items were more or less equally difficult to learn. That is, the material did not have any characteristics that were predictive of learning which could be used to discriminate between the items, but, because JOLs could likely rely on idiosyncratic cues that appear after studying, the JOLs were more accurate than the EOL judgments. It is likely that if participants had judged material varying in some predictive characteristic, the EOL judgments could have been more accurate, perhaps as accurate as the JOLs. Although it is possible JOLs would still be more accurate because they could rely both on idiosyncratic cues and item characteristics, the difference in accuracy would likely have been smaller. Investigating this issue was the one of the aims of Study II.

How do people make metacognitive judgments?

According to Nelson and Narens’ (1990) theoretical framework (Figure 1), people continuously monitor the object level, but how do people make metacognitive judgments? Although several theories about metacognitive monitoring exist (see, Nelson, Gerler, & Narens, 1984), it is fairly well-established that metacognitive monitoring is inferential (e.g., Koriat, 1997). In other
words, metacognitive judgments are made by making inferences about how available cues influence learning or memory.

The cue-utilization approach

Within the inferential view of metacognition, Koriat (1997) developed the cue-utilization framework. According to this framework people use available information in a learning situation as cues to make inferences about past, current, or future memory and learning. The cues include characteristics in the material, such as word length, how common words are, or the concreteness of words (Jönsson & Lindström, 2010). Other internal subjective cues include how easy an item is to retrieve from memory (e.g., Nelson & Dunlosky, 1991), the feeling of familiarity associated with an item (e.g., Reder & Ritter, 1992) or the experience of effort associated with processing an item (e.g., Begg et al., 1989). People may also use the current learning context as a cue, for example, when material is either read or listened to (Peynircioğlu et al., 2014). It is common that metacognitive judgments are less than perfectly accurate. According to the cue utilization framework, this discrepancy between memory and monitoring can be explained by which cues people use. If memory is influenced by a factor but people ignore it and instead rely on cues not predictive of memory, their judgments will be inaccurate, creating a discrepancy between judgments and memory.

Hence, the cue-utilization framework is consistent with the observation that memory and metacognition is dissociable. In other words, if an aspect of a learning situation affects memory it does not automatically affect metacognition. Thus, people may be unaware of aspects of a learning situation that affect their memory and may use cues to make inferences about their memory which, in fact, do not affect memory (e.g., Rhodes & Castel, 2008). Thereby, monitoring may be inaccurate if people ignore cues that are predictive of their memory, or if they rely on cues which do not affect memory. Evidence from laboratory research supports the view that memory and metacognition is dissociable (Koriat & Lieblich, 1977; Metcalfe, 2000; Metcalfe et al., 1993; Reder & Ritter, 1992; Schwartz & Smith, 1997). One type of strong evidence for the inferential view is metacognitive illusions. Several studies have demonstrated that people sometimes base their monitoring on information which is completely unrelated to memory (e.g., Rhodes & Castel, 2008, 2009), and ignore aspects which have an effect on memory (e.g., Sungkhasettee, Friedman, & Castel, 2011). For example, people judged words as being more likely to be remembered if the words were presented in a large font compared to a small font (e.g., Mueller et al., 2014; Rhodes & Castel, 2008), or written with their dominant hand compared to their non-dominant hand (Susser, Panitz, Buchin, & Mulligan, 2017), although neither the hand one writes with nor font size have an effect on memory. Consequently, evidence supports an inferential view of metacognition.
How do people use multiple cues?

Although the use of cues in metacognitive monitoring is supported by a considerable amount of evidence (e.g., Koriat, 1997), the details about cue-use are not yet well understood. For example, it is still unclear how many cues people use at the same time (or can use), how multiple cues are combined or weighted, and whether people prefer certain cues over others. Investigating how multiple cues are used simultaneously when people make metacognitive judgments is important in order to gain a deeper theoretical understanding of how metacognitive judgments are made. The importance of understanding how people combine multiple cues was recently noted by Rhodes (2016, p. 77), who wrote “it is essential to achieve an understanding of how multiple cues function collectively to inform judgment.”

In relation to the question of how many cues people can use at the same time, Undorf, Söllner and Bröder (2018) showed that some participants were able to integrate four cues at once. Specifically, JOLs were influenced by how many times a word had been studied, whether it was presented in large or small font, as well as the concreteness and emotionality of the word. Therefore, people seem to have the ability to integrate four cues into one judgment. However, whether or not this is the limit for how many cues people can use is unknown, and it is possible that people have the ability to use even more cues.

Investigating how people use multiple cues was one of the aims of Study II in this thesis. Specifically, in Study II, my colleagues and I investigated how multiple cues are combined, and propose that cues may compete for influence over metacognitive judgments. We refer to this as the cue competition hypothesis. Specifically, the level of influence a cue has on a judgment is determined by how salient the cue is and how predictive it seems to be of what is judged, in relation to other available cues. For example, it is possible that a salient and predictive cue such as semantic relatedness between words in a word pair can lower the influence of other less salient and predictive cues. A related idea, but for another judgment task, is what has been termed judgmental overshadowing (Price & Yates, 1993). Judgmental overshadowing is the phenomenon that, when people judge how predictive a cue is of an outcome variable, their judgments differ depending on how predictive other available cues are. For example, Price and Yates (1993) investigated this phenomenon by asking participants to predict the likelihood that a patient had the disease chronitis depending on whether or not the patient had fever/no fever or a rash/no rash. Before making the predictions, participants completed a series of learning trials during which they were presented with patients with combinations of the two symptoms and were informed of whether the patients had chronitis or not. Participants were divided into two groups. For both groups fever was always moderately predictive of chronitis, while the predictive power of a rash depended on the group. For one group, a patient having a rash was highly predictive of the disease, and for the other group a rash had low predictive power.
The results showed that participants judged the predictive power of fever differently in the two groups. Specifically, when rash had high predictive power, participants judged fever to be less predictive of chronitis than when rash had low predictive power. Hence, their results indicate, at least in a judgment and decision-making context, that how predictive a cue seem depends on how predictive other available cues are. If these findings generalize to judgments about one’s own memory and learning, it could presumably influence how people use cues in that context. In other words, if a salient and predictive cue makes a second cue seem less predictive, people may rely on the second cue less.

Related to this idea, there is research indicating that one cue can moderate how another cue influences metacognitive judgments. For instance, in one experiment, Rhodes and Castel (2008) found JOLs to be influenced by whether a word was presented in a large or small font size. However, when they added the salient and predictive cue of semantic relatedness, in another experiment, font size had a smaller impact on the JOLs. This indicates that semantic relatedness lowered the impact of font size.

In addition, studies investigating the combined use of feeling of familiarity and memory retrieval on JOLs indicate that retrieval can moderate the influence of familiarity. In one study, participants first studied picture-word pairs and then subsequently made JOLs with only the picture present (Metcalfe & Finn, 2008b). Moreover, participants made both speeded and unspeeded JOLs to investigate how familiarity affected JOLs, when participants did and did not have time to try to retrieve the word from memory. The results indicate monitoring occurred in two stages. In the unspeeded condition, participants first used the familiarity of the picture as a cue, and, if it the picture did not feel familiar, participants responded with a quick low JOL. Meanwhile, if familiarity was higher, participants made slower judgments, and the JOL was influenced by whether or not the word could be retrieved from memory. In the speeded condition, participants instead primarily relied on their feeling of familiarity. That is, when only familiarity was available as a cue (in the speeded condition), participants gave higher JOLs when familiarity was high, and lower when it was low. In contrast, when they could also rely on a retrieval attempt as a second cue (the unspeeded condition), JOLs were primarily influenced by the outcome of the retrieval attempt. In other words, familiarity was used differently as a cue when participants could make a retrieval attempt. It should also be noted that a similar pattern of results has been found when FOK judgments have been investigated (Koriat & Levy-Sadot, 2001).

In summary, people can use multiple cues when they make metacognitive judgments, but is unclear how multiple cues are used at once. Studies of judgments of how predictive a cue is, are affected by how predictive other available cues are (Price & Yates, 1993). If this cue-interaction effect generalizes to cues used to make judgments about one’s own learning and memory, it could influence how cues are used in that context. Furthermore, the idea that cues
compete for influence over metacognitive judgments is supported by some
evidence in the metacognitive literature, and was further investigated in Study
II.

All three studies in this thesis are related to cue-use in metacognitive mon-
itoring. Specifically, in Study I of this thesis, my colleagues and I investigated
how the accuracy of EOL judgments is affected by how much a predictive cue
varies between items. Namely, if there is more variation in a predictive cue,
this should allow participants to discriminate between easy and difficult items
than if the cue varies less. In Study II, this investigation was continued, but
also focused on how the accuracy of judgments changes from the beginning
of learning (i.e., EOL judgments) to later in learning (i.e., JOLs) because idi-
osyncratic cues become available during study. Study II also focuses on
whether cues may compete for influence over metacognitive judgments. Fi-
nally, in Study III, we investigated how processing fluency (described in the
next section) is used as a cue and how people’s metacognitive beliefs (i.e.,
metacognitive knowledge) guide how they use cues.

Processing fluency
One especially interesting cue found to affect metacognitive monitoring is
processing fluency, which has been described as the effort associated with
processing information (Alter & Oppenheimer, 2009). Processing fluency af-
fects many everyday judgments, such as when people judge the truth of a
statement (Hasher, Goldstein, & Toppino, 1977; Reber & Schwarz, 1999), or
when they judge the intelligence of an author (Oppenheimer, 2006). More im-
portantly for this thesis, processing fluency is likely to be frequently used as a
cue in metacognitive monitoring. For instance, Begg et al. (1989) proposed
that metacognitive judgments are based on processing fluency. For example,
a more common word (e.g. hungry) may be processed faster than a less com-
mon one (e.g., peckish), and people will use how fast or easy an item is pro-
cessed as a cue for how easy it will be to learn or remember. Although Begg
and his colleagues did not measure processing fluency, later studies have
shown that JOLs are negatively correlated with how fast people process an
item (Hertzog et al., 2003). That is, an item that is processed fast is given a
higher JOL. This indicates that processing fluency is an influential cue used
for metacognitive judgments.

Processing fluency can be divided into several distinct sub-categories
(Alter & Oppenheimer, 2009). However, not all of them are relevant for the
purpose of the current thesis, and therefore I will only describe two of the
categories. One type of fluency is perceptual fluency, that is, how fluently
people can perceive a stimulus. This has been manipulated by changing per-
ceptual aspects of a stimulus to make them easier or more difficult to perceive.
For example, perceptual fluency is manipulated by presenting words in alter-
nating upper- and lower-case letters (e.g., sTiMuLuS) or only lower-case letters (e.g., stimulus; Rhodes & Castel, 2008). Perceptual fluency may also be manipulated by presenting words in a clear font or an unclear font (e.g., Simmons & Nelson, 2006). Memory-based fluency is another sub-category of processing fluency. One type of memory-based fluency is retrieval fluency, that is, how easily or fast information is retrieved from memory (e.g., Benjamin, Bjork, & Schwartz, 1998). For example, research has shown people respond with higher JOLs to answers that are quickly retrieved from memory than answers that take longer to retrieve from memory (e.g., Koriat & Ma’ayan, 2005). Furthermore, memory-based fluency also includes encoding fluency, that is, the feeling of ease or speed by which information is encoded to memory (e.g., Hertzog et al., 2003). Although processing fluency likely affects many types of judgments, some research suggested that people’s beliefs about fluency may mediate the effect fluency has on metacognitive monitoring (Mueller & Dunlosky, 2017).

Metacognitive beliefs

Flavell (1979) described that people’s metacognitive knowledge, that is, beliefs and knowledge about cognition, play an important role in metacognitive monitoring. Hence, if people believes that, for example, high fluency (e.g., an item feels effortless to encode to memory) indicates items are easy to learn, they may judge high-fluency items as being easier to learn than low-fluency items. In contrast, if they instead have the opposite belief, they may instead judge high-fluency items as being less likely to be learned. There is some evidence that for JOLs the effect of font size (Mueller et al., 2014) and semantic relatedness between words in a word pair (Mueller, Tauber, & Dunlosky, 2013), are mediated by people’s metacognitive beliefs. As mentioned earlier, Rhodes and Castel (2008) showed that people tend to give higher JOLs to words presented in a large font than a small font, and the authors argued that this font-size effect may be the result of processing fluency, as large font words may feel like they are more easily processed. However, later research indicates that it could instead be people’s beliefs about the font size that guide their JOLs (Mueller et al., 2014). Thus, if people believe large words are easier to remember than small words, their beliefs will guide their JOLs. Mueller et al. (2014) investigated this by asking participants to make so-called pre-study JOLs for words and by measuring processing fluency. Pre-study JOLs differ from regular JOLs, because they are made without knowing what word is about to be studied. Instead, before each word was studied, participants were told that the upcoming word would either be a large or a small word, and they then judged how likely it would be that they would be able to remember the word on a later test. Their results showed pre-study JOLs were affected by font size in a similar way as JOLs made while viewing the words. In addition,
the results indicated that font size did not impact processing fluency. Consequently, the authors argued that beliefs about how font size affects memory mediated the font size effect. Hence, beliefs likely have an impact on people’s use cues when they make metacognitive judgments. However, there is also research indicating that processing fluency may influence metacognitive judgments irrespective of beliefs. In one study, researchers manipulated the loudness by which items were presented to participants. The loudness was argued to affect processing fluency, as lower volume would be associated with lower processing fluency. In the study, participants gave higher JOLs to items presented in a loud volume compared to a quieter volume, and some participants did this even when it contradicted their beliefs (Frank & Kuhlmann, 2017). Furthermore, when semantic relatedness and font size is manipulated, both accuracy and magnitude of immediate JOLs differ from that of pre-study JOLs (Price & Harrison, 2017), indicating there are factors other than beliefs that affect metacognitive judgments, and processing fluency could be such a factor. Hence, it is possible that processing fluency affects judgments despite how people believe a manipulation affects learning or memory.

It is also possible that processing fluency and beliefs interact in metacognitive monitoring. In their analytic processing theory, Mueller and Dunlosky (2017) suggested that beliefs have a central role in determining how cues influence JOLs. According to the theory, asking people to make memory predictions triggers an analytic problem-solving mindset. In this mindset, people start looking for cues that can be used to make the asked for predictions. Also, if people do not have any pre-existing beliefs about the cues, people may create beliefs based on the current learning situation. For instance, when words are presented in different font sizes, participants may form beliefs as to how this manipulation will affect their memory (e.g., Mueller et al., 2014). However, the theory does not eliminate the possibility of the influence of processing fluency on monitoring. Instead, the authors argue that if people do not have beliefs about processing fluency, it may influence monitoring without beliefs.

Although a multitude of cues have been shown to affect EOL judgments, it is still unclear exactly how they influence them. First, we know little about the relative impact of multiple cues on EOL judgments, which is further complicated by the fact that experimentally manipulated cues may be inter-correlated. Second, a cue may influence an EOL judgment both because of the participants’ explicit and analytic beliefs about how that cue relates to learning, but also in a more automatic fashion through a nonconscious inference. That is, perhaps participants intuitively sense how easy an item is to learn without having an explicit idea of why.

Whether or not people actively apply their beliefs about the different characteristics to make EOL judgments, or if they are influenced by how easily they can process an item, has never been tested. However, it has been argued
that, for example, word frequency should affect EOL judgments through processing fluency (Begg et al., 1989). This relates to the second major aim of this thesis – what information people use to make EOL judgments. Investigating whether processing fluency affects EOL judgments, and whether its influence is mediated about beliefs, was investigated in Study III.

The aims of this thesis

Although EOL judgments have been investigated since the 1950s, the literature is still fairly limited, and what affects EOL judgments and their accuracy has not yet been systematically investigated (as also noted by other authors, Dunlosky & Metcalfe, 2009). To understand how people make EOL judgments, it is central to understand what types of information they rely on as cues, as well as how these cues affect judgment accuracy. In order to better understand EOL judgments it would also be fruitful to compare the judgments with the similar and much more widely investigated JOLs. By identifying similarities and differences between the two types of judgments, it will aid the theoretical work of describing EOL judgments. When people make EOL judgments, they presumably rely on both their knowledge about what is easy or difficult to learn, but they may also rely on more subtle experiences, such as, for example, the experience of processing fluency. Related to these unresolved issues in research of EOL judgments, the studies included in this thesis had two major aims with related research questions.

The first aim was to investigate how accurate EOL judgments are and what moderates their accuracy.

(i) How does the discriminability of a predictive cue affect EOL judgment accuracy (Study I and II)?
(ii) With near optimal circumstances, how accurate are EOL judgments (Study I)?
(iii) Does experimental methodology influence EOL judgment accuracy (Study I)?
(iv) Can EOL judgments be as accurate as JOLs (Study II)?

The second aim was to investigate what cues people use when they make EOL judgments and how people use cues to make metacognitive judgments.

(v) Do cues compete for influence (Study II)?
(vi) Is processing fluency used as a cue for EOL judgments (Study III)?
(vii) How do people’s beliefs about a cue impact how it is used (Study III)?
Empirical studies

Study I: Jemstedt, Kubik, and Jönsson (2017)

The overall aim of Study I was to identify potential moderators of EOL judgment accuracy and how well people can predict future learning when the judged material varies in a predicative cue. Because the literature on EOL judgments is fairly limited and a wide range of procedures have been used to measure EOL judgments, we also wanted to investigate methodological aspects that could potentially influence accuracy. Understanding the possible influence of experimental methodology on EOL judgment accuracy is important both for understanding EOL judgments in general, but also for gaining an understanding of which methods to use in future studies.

Experiment 1

Question I - How does the discriminability of a predictive cue affect EOL judgment accuracy?

Accuracy of EOL judgments varies between studies (e.g., Jönsson & Kerimi, 2011; Leonesio & Nelson, 1990). One possible reason for this may be the difference in the material composition used in the studies. More specifically, with higher variation in the predictive characteristics of items, participants may better discriminate between easy and difficult items and accuracy will improve. However, this has not previously been directly manipulated with EOL judgments. Therefore, our first aim was to investigate whether accuracy of EOL judgments is moderated by how much an available predictive cue varies in the judged material. We predicted that EOL judgment accuracy would be higher when the material had high variability than when the material varied less.

Question II - With near optimal circumstances, how accurate are EOL judgments?

Previous studies have shown that EOL judgments can be accurate (e.g., Jönsson & Kerimi, 2011; Underwood, 1966), but the accuracy has differed greatly between studies. Therefore, we wanted to investigate how accurate EOL judgments could be given near-optimal circumstances. Specifically, we wanted to
investigate how high accuracy would be when participants had access to a highly predictive cue which varied considerably between items. We predicted that we would find EOL judgments to be highly accurate in such a condition.

**Question III - Does experimental methodology influence EOL judgment accuracy?**

According to the notion of transfer-appropriate monitoring (e.g., Dunlosky & Nelson, 1992), monitoring accuracy will be higher the more similar the judgment task is to the criterion measure. Therefore, we predicted that accuracy would be higher when the EOL judgment task matched the criterion measure compared to when there was less of a match. Specifically, the aim was to investigate to what extent the judgment task and criterion measure would moderate the accuracy, and whether the task and criterion would interact to influence accuracy. To investigate this, we used two judgment tasks and two criterion measures. In one task, participants made EOL judgments by selecting how easy or difficult each item would be to learn, and, in the other task, they selected how many study trials they would need to learn each item. Following the judgment phase, the participants completed three consecutive study and test phases. After the experiment, judgments were correlated with whether or not each item had been learned on the first test, or how many study trials were needed to learn each item. We predicted that, when judgments in the trials-to-learn EOL task were correlated with the trials-to-learn criterion, EOL accuracy would be higher than in the other combinations of judgment and criterion. Furthermore, we also aimed to investigate whether the criterion measure would have an effect on accuracy itself. A binary criterion measure could potentially reduce accuracy because of restriction of range (Schwartz & Metcalfe, 1994). Using a trials-to-learn criterion could therefore potentially minimize this problem, and better measure a real, higher, accuracy.

**Methods**

Thirty-one psychology students at Stockholm University participated in the experiment. We used a design in which we manipulated three variables: (1) how much the word pairs varied in semantic relatedness, (2) whether EOL judgments were made by selecting how difficult a pair would be to learn or how many study trials had to be completed to learn the pair (trials-to-learn; TTL), and (3) whether the criterion measure was the outcome on the first test (binary) or how many study trials were completed before a pair was correctly recalled on a test (trials-to-learn; TTL). Specifically, we used a 2 (list: high vs. low variability) × 2 (EOL task: difficulty vs. TTL) × 2 (criterion: binary vs. TTL) within-group design.

Each participant completed two blocks with an identical procedure. In one block participants judged, studied and were tested on a list of word pairs varying highly in semantic relatedness (high variability list), that is, word pairs varied from very low relatedness (e.g., ndege–fuvu) to high relatedness (e.g.,
In the other block, all pairs had low semantic relatedness (low variability list; e.g., baby–north). The order of the blocks was counterbalanced between participants.

In each block, participants first completed two types of self-paced EOL tasks, first one, and then the other. Both tasks had the same EOL judgment question, namely “How difficult will it be to learn the association between the right and left word?” First, they completed the difficulty task, in which they judged the difficulty of the items on a scale from “0% (1)” (most difficult to learn), “20% (2),” “40% (3),” “60% (4),” “80% (5),” or “100% (6)” (easiest to learn) by pressing the corresponding button on the keyboard (e.g., 2 for 20%). In the second, trials-to-learn (TTL) task, participants instead selected how many study trials they would need to learn the item by selecting “1 = I will have learned it by Test 1,” “2 = I will have learned it by Test 2,” “3 = I will have learned it by Test 3,” or “4 = I won’t have learned it by Test 3.” Following the EOL judgment phase, three study and test phases were completed. During each study phase, each item was studied for 5 seconds. After each study phase, a cued-recall test for all items was completed. During the test participants had 15 seconds to recall the correct target-word when presented with each cue-word (e.g., baby – ?). When the first block had been completed, the next block followed seamlessly. The order of all items was fully randomized in all phases of the experiment. In addition, between each phase of the experiment, participants completed a 45-second distraction task. After the experiment, judgment accuracy was calculated with a Goodman-Kruskal gamma correlation between the judgments from the two EOL judgment tasks and two criterion measures. The binary criterion consisting of recall (correct or not correct) on the first test, and a TTL criterion consisting of which test each item was first recalled on (or if it was not recalled on any test).

**Main results**

As can be seen in Figure 3, there was a large main effect of list. That is, participants made more accurate judgments when the items in the list varied highly in semantic relatedness ($M = .74, SD = .13$) compared to when the variability was low ($M = .21, SD = .24$). Hence, because semantic relatedness is predictive of learning, and the participants could use the cue to discriminate between items, their judgments were more accurate for the high variability list. The results also showed that judgments in the TTL-task ($M = .50, SD = .15$) were more accurate than judgments in the difficulty-task ($M = .45, SD = .16$), indicating that participants were slightly better at predicting their learning when asked to judge how many study trials each item would take to learn. Finally, there was also a small but reliable main effect of criterion measure, indicating that the correlations were higher with the binary criterion ($M = .484, SD = .13$) compared to the TTL-criterion ($M = .465, SD = .14$). We predicted
that correlations would be higher when judgments in the TTL-task was correlated with the TTL-criterion, but the interaction was not reliable.

Figure 3. EOL accuracy (G), as a function of list (high vs. low variability) and EOL task (difficulty vs. TTL), separately shown for the two criterion measures, binary (first panel) and TTL (second panel). Error bars represent the 95% confidence intervals.

Conclusion
When people make EOL judgments, they rely on the available cues in the material. As predicted, when the judged list varied in semantic relatedness, the accuracy was higher than when the list had low variability of semantic relatedness. That is, without any variation in a predictive cue, EOL judgments are fairly inaccurate, but with variation in the cue it allowed participants to discriminate between items, and consequently the EOL judgments were accurate. Moreover, the results indicate that using a TTL-task to make EOL judgments may make them more accurate than using a difficulty-task (which is the most common type of task in the literature.) However, because of a mistake in the programming of the experiment, the TTL-task was, in Experiment 1, always the second judgment for each item, and therefore it is also possible that it was the order of the judgment rather than the task that affected the accuracy. This possibility was tested in Experiments 2 and 3. Finally, the binary criterion measure produced slightly higher correlations than the TTL-criterion. We did not predict this, but one possible explanation is what is called test-potentiated learning (Arnold & McDermott, 2013; Kubík, Olofsson, Nilsson, & Jönsson, 2016). Namely, after a test, test performance may make people aware that some items are better learned than others, and this can influence how they study the items after the test. Because the binary criterion was based on only the first test, it was not affected by test-potentiated learning, whereas the TTL-test was based on all three tests and thus would be affected by it. In other words, the idiosyncratic learning that follows the first test is likely hard to
predict in the beginning of learning, and therefore judgments may better predict performance on the first test than all three tests.

Experiment 2

**Question III - Does experimental methodology influence EOL judgment accuracy?**

The aim of Experiment 2 was to further investigate whether the judgment task (difficulty vs. TTL) would have an effect on EOL accuracy.

**Methods**

Forty-eight psychology students were divided into two groups. Both groups completed the same experimental procedure, but one group made EOL judgments in a difficulty-task and one group in the TTL-task. Both groups judged, studied, and were tested on the high-variability list used in Experiment 1. The experiment started with participants making EOL judgments for 8 seconds per item, then studying the word pairs for 5 seconds. After which they completed a 1-minute distractor task, and then finally a cued-recall test. All participants were told in the beginning of the experiment that there would be three study and test trials, but the experiment was stopped after the first test. Hence, judgments were only correlated with test performance on one test.

**Main results**

There was no reliable difference in EOL accuracy depending on what EOL task participants made their judgments in (difficulty task: $M = .82$, $SD = .15$; TTL task: $M = .76$, $SD = .15$).

**Conclusion**

The results of Experiment 2 indicate that the higher EOL accuracy in the TTL-task found in Experiment 1 may not be the result of the EOL task itself. Instead, it may be a result of the order of the EOL tasks or perhaps a Type 1 error, that is, a false positive brought about by chance.

Experiment 3

**Question III - Does experimental methodology influence EOL judgment accuracy?**

Metacognitive judgments are highly comparative (Koriat, 1997), that is, the difficulty of one item is judged in comparison to other items. Moreover, when participants in a study were given a low learning goal (learn a low number of items), and asked to select which items to study, their study decisions differ depending on how the items are presented (Dunlosky & Thiede, 2004). When
all items were presented at the same time, participants focused on easy items, but when items were presented item-by-item they instead focused on difficult items (Dunlosky & Thiede, 2004). That is, presentation method affected study decision. When the learning goal is low, the best study strategy is to focus on easier times, but participants primarily did this when they could see all items at once. One possibility for this difference is that participants could more accurately judge which items were easy and which were difficult when they could see all items at once. However, this has not been tested.

Therefore, in Experiment 3, we aimed to investigate whether EOL judgment accuracy is moderated by presentation method during the judgments. More precisely, if accuracy differs when EOL judgments are made sequentially (i.e., item-by-item) or simultaneously; that is, with the full list of items visible when each judgment is made. One reason EOL judgment accuracy could be better when judgments are made with all items visible is because accuracy depends on how well people can predict whether one item will be learned compared to other items. If EOL judgments are made item-by-item, each judgment may only be compared to items that come to mind during each judgment, whereas, when all items are visible, each item can be compared to all other items.

We also further investigated whether accuracy would differ when judgments were made using a TTL-task or a difficulty-task, and when judgments were correlated with a binary criterion or a TTL-criterion.

Methods
Sixty-four psychology students were divided into two groups and made judgments, studied, and were tested on a list of word pairs varying in semantic relatedness (similar to the high variability list in Experiment 1). We used a 2 (presentation format: simultaneous vs. sequential) × 2 (EOL task: difficulty vs. TTL) × 2 (criterion: binary vs. TTL) mixed design. EOL task was manipulated between group, that is, one group made EOL judgments in a TTL-task and the other in the difficulty-task. Presentation format and criterion measure were manipulated within-group. Presentation format was manipulated within-group by dividing the experimental procedure into two blocks; in one block judgments were made sequentially, and in the other simultaneously. After the experiment, judgments were correlated with both the binary and the TTL criterion.

The experimental procedure in each block was similar to Experiment 1, that is, first judgments were made, after which three consecutive study and test trials were completed. After the first block was completed the second block continued seamlessly. A 45-second distractor task was completed between each phase of the experiment.
Main results
The results did not support our hypothesis that EOL accuracy would be higher when items are presented simultaneously ($M = .60$, $SD = .28$) compared to sequentially ($M = .63$, $SD = .18$). That is, presentation formation had no reliable effect on EOL judgment accuracy.

In addition, there were no reliable difference in accuracy between judgments made in the TTL task ($M = .62$, $SD = .20$) and the difficulty task ($M = .61$, $SD = .19$). However, we replicated the main effect of criterion measure found in Experiment 1. Correlating judgments with the binary criterion ($M = .64$, $SD = .20$) once again resulted in higher correlations than correlating judgments with the TTL-criterion ($M = .59$, $SD = .19$).

Conclusion
We reasoned that, because of the comparative nature of metacognitive judgments (Koriat, 1997), EOL judgment accuracy would be higher when all items were presented simultaneously during the EOL judgment phase instead of sequentially. However, our results did not support this idea and instead indicated that perhaps people do not need access to all other items to make accurate EOL judgments. Instead, it is possible people make EOL judgments relative to learning in general, or maybe they continuously update an average item difficulty to which they compare other items.

We also further investigated the influence of the EOL task and criterion measure on accuracy of EOL judgments. Together with the results from Experiment 2, our results indicate that making EOL judgments in a TTL-task does not lead to more accurate EOL judgments than in a difficulty-task. That is, superior accuracy of the TTL-task in Experiment 1 most likely resulted from some other factor than the task itself, possibly the order of the task.

In Experiment 3, we also replicated the higher correlations using the binary criterion compared to the TTL-criterion. For a possible explanation to this result, see the conclusion of Experiment 1.

Study II: Jönsson, Jemstedt, Schwartz, and Kubik (2017)
In Study II we continued our investigation of EOL judgment accuracy. Study II explored questions related to both of the major aims of this thesis. In so doing, we wanted to investigate whether EOL judgments could be as accurate as JOLs given that the material varied with respect to predictive item characteristics. In addition, we wanted to investigate whether cues compete for influence over metacognitive judgments. That is, we were interested in whether people will rely less on one cue when another cue is available as the cue landscape changes.
It is important to understand which, and how, cues influence metacognitive accuracy because accuracy likely determines how effectively participants control their study when they learn (Nelson & Narens, 1990). As mentioned in the introduction, an earlier study concluded that JOLs are more accurate than EOL judgments (Leonesio & Nelson, 1990). However, as we showed in Study I, EOL judgments can be highly accurate, possibly as accurate as JOLs. It is possible that Leonesio and Nelson (1990) found that EOL judgments were less accurate than JOLs because they used a homogenous material. During study the cue landscape changes, and JOLs may rely on idiosyncratic cues, which signal how well each item had been learned during study. This shifting of cues then leads to JOLs being more accurate than EOL judgments. However, if the item characteristics are more varied, then EOL judgments may be accurate as well. Nonetheless, even when a predictive item characteristic varies, JOLs still have the benefit of being able to rely on idiosyncratic cues, and it is possible they may still be more accurate than EOL judgments. Still, the difference in accuracy would likely be smaller when items vary in a predictive item characteristic.

Experiment 1

Question IV - Can EOL judgments be as accurate as JOLs
The aim of Study II was to compare the accuracy of EOL judgments with JOLs when the variability of a predictive item characteristic is manipulated (as in Study I). That is, we predicted that JOLs would be more accurate than EOL judgments when word pairs did not vary in semantic relatedness, but that there would be no reliable, or a smaller, difference in accuracy between EOL judgments and JOLs when the word pairs varied in semantic relatedness. Showing this difference will contribute to our theoretical understanding of EOL judgments by demonstrating that they may be as accurate as JOLs when there is sufficient information available in the environment.

Question V - Do cues compete for influence?
Throughout the learning process the cue landscape changes. Different cues are available at different times, and some cues become less relevant and salient, depending on when a judgment is made and what people are judging. The main difference in the cue landscape between EOL judgments and JOLs is that, after studying a material, people may gain access to additional idiosyncratic cues, which allow them to make more accurate JOLs. However, it is possible the influence of these idiosyncratic cues on JOLs is limited if another more salient and predictive cue is available. More precisely, perhaps idiosyncratic cues impact JOLs to a smaller degree when a set of word pairs varies highly in semantic relatedness than when it varies less. In other words, cues may compete for influence over metacognitive judgments. We refer to this
hypothesis as the cue competition hypothesis. In line with this hypothesis, we predict that JOLs will be more accurate than EOL judgments when semantic relatedness has low variability. Meanwhile, when semantic relatedness varies considerably between items, the difference should disappear, or at least be smaller.

Methods
Sixty-six psychology students at Stockholm University were recruited and randomized into two groups. To investigate our hypotheses, we used a mixed design in which half of the participants judged, studied, and were tested on a high-variability material, and for the other participants a low-variability material (similar to the materials in Study I). Judgment type was manipulated within-group, as EOL judgments were completed for half the items in each material and JOLs for the other half.

Participants first made self-paced EOL judgments for half of the items, then studied all items for 5 seconds each. After studying all the items, self-paced JOLs were made for the half of the items that were not used in the EOL judgment phase. When JOLs were made, both cue and target words were visible (cue-target JOLs). Finally, after making the cue-target JOLs, a cued-recall test was completed for all items. During the test, participants had a maximum of 15 seconds to respond to each item. Throughout the experiment, presentation order of all the items was randomized. Moreover, between each phase of the experiment, a 30-second distraction task was completed.

Main results
Because the accuracy measures (i.e., gamma correlations) were not normally distributed, we analyzed median accuracy instead of mean accuracy.

In line with the cue competition hypothesis, we predicted that cue-target JOLs would be more accurate than EOL judgments for the low-variability list, but that there would be a smaller, or no, difference for the high-variability list. As can be seen in Table 1, cue-target JOLs were more accurate than EOL judgments for the low-variability list, but there was no reliable difference for the high-variability list. That is, the results support the cue competition hypothesis.
Table 1. Median Goodman-Kruskal gamma correlations for the EOL judgments and cue-target JOLs (Exp. 1 and 2), and cue-only JOLs (Exp. 2) as a function of material (high- vs. low-variability).

<table>
<thead>
<tr>
<th>Material</th>
<th>Judgment type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EOL judgment</td>
</tr>
<tr>
<td><strong>Experiment 1</strong></td>
<td></td>
</tr>
<tr>
<td>High-variability</td>
<td>.58*</td>
</tr>
<tr>
<td>Low-variability</td>
<td>.11</td>
</tr>
<tr>
<td><strong>Experiment 2</strong></td>
<td></td>
</tr>
<tr>
<td>High-variability</td>
<td>.51*</td>
</tr>
<tr>
<td>Low-variability</td>
<td>.10</td>
</tr>
</tbody>
</table>

* The asterisk shows that monitoring accuracy is significantly different from zero, that is above chance ($p_s < .001$), as tested with a Wilcoxon one-sample signed rank test. No asterisk: $p_s > .05$.

Conclusion
The results in Experiment 1 showed EOL judgments can be as accurate as cue-target JOLs when the material varies in a salient characteristic predictive of performance on the memory test. However, when the material did not vary in the predictive cue, cue-target JOLs were more accurate than EOL judgments. That is, we replicated older research showing that cue-target JOLs are more accurate than EOL judgments (Leonesio & Nelson, 1990). Nevertheless, we also showed that this finding has limited generalizability and depends on what material is used.

The results also supported the cue competition hypothesis. When a salient and predictive cue is available, it can compete with other cues and make them less influential. Although idiosyncratic cues should have been available for the JOLs, this did not increase their accuracy above that of EOL judgments. However, it is also possible JOLs made on the high variability list actually relied on both the characteristics in the material, as well as idiosyncratic cues, but both types of cues indicated the same items as being more likely to be remembered later. In other words, perhaps word pairs with high semantic relatedness were also the ones that felt better learned, and thereby the idiosyncratic cues did not have room to improve the accuracy of the JOLs. This possibility is further discussed in the general discussion.
Experiment 2

Questions IV and V

The aim in Experiment 2 was the same as in Experiment 1, but we added a third judgment-type condition, in which participants made cue-only JOLs. That is, participants either made EOL judgments, JOLs with both cue and target words visible (cue-target JOLs), or JOLs with only the cue-words visible (cue-only JOLs). In addition to the prediction in Experiment 1, we also predicted that cue-only JOLs would be more accurate than the other types of judgments, independent of the material. We predicted this because research has shown that cue-only JOLs can be highly accurate because people can rely on the outcome of a retrieval attempt (e.g., Nelson & Dunlosky, 1991). By this, I mean that, when they are presented with only the cue-word they will try to retrieve the target word from memory. Whether or not they are able to remember the target word is highly predictive of retrieval of the same item on the later memory test. In line with the cue competition hypothesis, we also predicted that cue-only JOLs would rely mainly on retrieval attempts rather than semantic relatedness, partly because participants would only be able to use semantic relatedness for the words they were able to recall, but also because outcome of the retrieval attempt would be a more salient and predictive cue. To further investigate how semantic relatedness was used in the three judgment type conditions, we compared correlations between pre-standardized semantic relatedness of the word pairs with the judgments in the different conditions.

Methods

We used a similar design to Experiment 1, but added a cue-only JOL condition, and all conditions were manipulated between groups. In this study, 274 psychology students were randomized into one of six groups, in which they either judged, studied, and were tested on a high- or low-variability list, and either made EOL judgments, cue-target JOLs or cue-only JOLs.

In the EOL judgment conditions, the experiment began with participants judging all the word pairs, and then studying the words for 5 seconds each and completing a cued-recall test. In the two JOL conditions, participants first studied the word pairs and then made their judgments before completing the cued-recall test. All item presentation was randomized, and participants completed a 30-second distraction task between each phase of the experiment.

Main results

We replicated the finding of Experiment 1 that cue-target JOLs were reliably more accurate than EOL judgments for the low-variability list, but not the high variability list. Again, this result indicates EOL judgments can be as accurate at cue-target JOLs when the material includes predictive cues. Furthermore,
we showed cue-only JOLs were more accurate than EOL judgments for both the high- and the low-variability list. Because participants could rely on a retrieval attempt in the cue-only JOL condition, their judgments become more accurate than EOL judgments, independent of material composition. Furthermore, cue-only JOLs were more accurate than cue-target JOLs for both the high- and the low-variability list. This replicates prior research showing that cue-only JOLs have superior accuracy compared to cue-target JOLs (e.g., Rhodes & Tauber, 2011), see Table 1 for median accuracy in each condition.

In addition, we were interested in which cues participants relied upon when they made judgments in the three judgment conditions. To investigate this, we correlated all judgments made on the high variability list, with the pre-standardized semantic relatedness of the word pairs. The median correlations in the three conditions clearly show that participants relied on semantic relatedness less when they made cue-only JOLs (.33) than when they made cue-target JOLs (.70), and EOL judgments (.75). Although EOL judgments relied on semantic relatedness slightly more than cue-target JOLs, this difference was not reliable.

Conclusion
The results show that when a material varies in a predictive cue, EOL judgments can be as accurate as cue-target JOLs. However, because JOLs are made after study, and some of the material has already been learned, people may rely on a retrieval attempt when the JOLs are made on the cue-word. Under these conditions, JOLs become more accurate than EOL judgments, independent of the material. As a result, JOLs will be more accurate than EOL judgments as long as people make JOLs in combination with a retrieval attempt. Nevertheless, it is likely possible to create material that varies in a highly predictive cue to such a level that EOL judgments will be as accurate as cue-only JOLs. However, such material is likely uncommon in everyday life. Furthermore, by comparing the median correlations between semantic relatedness of the word pairs and the judgments in the three conditions, it was clear that cue-only JOLs relied much less on semantic relatedness than the other two conditions. The same correlations for EOL judgments and cue-target JOLs did not differ reliably. A possible reason as to why there was no difference may be because the cue-target JOLs also relied on semantic relatedness. However, as previously mentioned, it could also be because semantic relatedness and idiosyncratic cues indicated the same items as being more likely to be remembered on the test. This could also be the reason why the cue-only JOLs were correlated with semantic relatedness. It is possible that participants did not rely on semantic relatedness for the cue-only JOLs, but word pairs with higher semantic relatedness were more likely to be retrieved from memory during the judgments. Consequently, this increased the correlation between the judgments and the semantic relatedness of the pairs. Unraveling these possibilities may
be an interesting avenue of research in order to better understand how cues interact to influence metacognitive judgments.

**Study III: Jemstedt, Schwartz, and Jönsson (2017)**

In Study I and II, we focused on EOL judgment accuracy. In Study III, we instead focused on the magnitude of EOL judgments and how it is affected by processing fluency and people’s beliefs about cues (i.e., metacognitive knowledge). Processing fluency has been identified as an important cue for JOLs (e.g., Frank & Kuhlmann, 2017; Rhodes & Castel, 2008), but has never been directly manipulated to investigate how it influences EOL judgments. In one study, processing fluency was indirectly manipulated by having participants make EOL judgments on words with different word frequency, that is, how often a word occurs in the English language (Begg et al., 1989). The authors argued that participants relied on processing fluency when they judged the words. They reasoned that words with higher word frequency were processed faster or easier and were therefore judged as easier, and vice versa. Thus, there is some evidence EOL judgments may be affected by processing fluency as well. However, Begg et al. (1989) did not directly manipulate processing fluency, and people may have relied on some other characteristics associated with the word frequency. Moreover, it has been argued that, when people rely on processing fluency, the effect of the fluency could be mediated by participants’ beliefs about how processing fluency should affect memory (e.g., Mueller & Dunlosky, 2017). Therefore, the overall aim of Study III was to investigate how processing fluency and beliefs affect EOL judgments.

**Experiment 1 and 2**

**Question VI – Is processing fluency used as a cue for EOL judgments?**

As mentioned above, whether people rely on processing fluency when they make EOL judgments has never been directly tested. It is likely that they do because processing fluency has been identified to affect JOLs (e.g., Frank & Kuhlmann, 2017; Rhodes & Castel, 2008). We predict that participants will judge items in a condition associated with high processing fluency as easier to learn than items in a condition associated with low fluency. In other words, people may be monitoring how effortful an item is to process and base their EOL judgments on this.

**Question VII - How do people’s beliefs about a cue impact how it is used?**

It has previously been showed the influence of some cues on JOLs are mediated by people’s beliefs about the cues (Mueller et al., 2014), and that beliefs
could also account for the influence of processing fluency (Mueller & Dunlosky, 2017). In contrast, other studies have indicated that beliefs about a cue possibly moderate the influence of processing fluency (e.g., Frank & Kuhlmann, 2017). Whether or not the effect of processing fluency on EOL judgments is mediated by beliefs has not been tested. Hence, it is both possible that processing fluency could influence EOL judgments because people believe it should influence learning, but also because processing fluency influence judgments independent of what people believe.

Methods
Both Experiments 1 and 2 used an identical procedure, but, because of a problem with the stimulus words used in Experiment 1, the methods below are for Experiment 2. I will briefly explain the problem with the words in Experiment 1 at the end of the methods section.

Participants were 146 students at Florida International University. All students made EOL judgments, studied, and were tested on six-letter concrete nouns with similar word frequency. We manipulated processing fluency within-group by randomizing words into two presentation condition associated with either high or low processing fluency. In the alternating condition, we produced low processing fluency by presenting words in a manner in which the letters of the words alternated between upper- and lower-case (e.g., bUcKeT). In the constant condition, we produced relatively higher processing fluency, by presenting words in only upper-case letters (e.g., BUCKET). Beliefs were not manipulated, but instead measured at the end of the experiment by asking participants to select which condition they believed made the words more likely to be learned. Participants could select between, “Only upper-case letters”, “Alternating upper- and lower case letters”, or “No difference”. Based on their response, participants were divided into an alternating, constant, or same belief group.

The procedure started with participants making EOL judgments for each word. Following this, they studied the words for 5 seconds each, after which they completed a 30-second distract task. After the distraction task, they completed a 2-minute free-recall test of all the words. Following the test, participants reported which condition they believed made the words more likely to be learned.

In Experiment 1, some of the words included the letters “I” and “L”, and when we scored the responses from the recall test, we noticed that participants mistook some of these words for other words when they were presented in the alternating condition (e.g., responding cattie instead of castle). This also affected how easy the words were to learn, as participants recalled fewer of the words from the alternating condition than the constant condition. Therefore, we repeated the same procedure in Experiment 2, but replaced all words including the letters “I” and “L”.
Main results

Because of the problem with some of the words in Experiment 1, I only describe the results of Experiment 2. However, it should be noted that, overall, the results of both experiments were nearly identical.

As can be seen in Figure 4, words presented in the alternating condition ($M = 43.76, SD = 19.55$) were judged as less likely to be learned than words in the constant condition ($M = 59.95, SD = 23.09$). However, an interaction between presentation condition and belief group indicated that the size of the difference between the two presentation conditions was moderated by participants’ beliefs. That is, the size of the difference was smallest when participants believed the alternating conditions made the words more likely to be learned, followed by the same condition, and then the constant condition. To be precise, contrary to their beliefs, participants in the same-group still judged words in the alternating condition as being less likely to be learned. Furthermore, in Experiment 2, presentation condition did not have a reliable effect on recall performance, that is, participants exhibited a metacognitive illusion. They judged words in the two presentation conditions would differ in how likely they would be to learn, but the presentation conditions had no effect on recall performance.
In line with research on JOLs (e.g., Frank & Kuhlmann, 2017; Rhodes & Castel, 2008), Study III showed processing fluency influenced metacognitive judgments. That is, we showed EOL judgments are affected by processing fluency as well. In addition, we showed that participants’ beliefs moderated the effect of processing fluency. It has been argued that beliefs can mediate the effect of fluency for JOLs (Mueller & Dunlosky, 2017), and our results indicate that this may not be the case for EOL judgments. Instead, people may hold a belief but make EOL judgments that contradict that belief. However, because we asked the participants about their beliefs at the end of the experiment, it is possible that they had other beliefs while making the EOL judg-

Figure 4. Mean EOL judgments and 95% confidence intervals factored by presentation condition and belief group. Number of participants in each belief group is presented below the group names.

Conclusions

In line with research on JOLs (e.g., Frank & Kuhlmann, 2017; Rhodes & Castel, 2008), Study III showed processing fluency influenced metacognitive judgments. That is, we showed EOL judgments are affected by processing fluency as well. In addition, we showed that participants’ beliefs moderated the effect of processing fluency. It has been argued that beliefs can mediate the effect of fluency for JOLs (Mueller & Dunlosky, 2017), and our results indicate that this may not be the case for EOL judgments. Instead, people may hold a belief but make EOL judgments that contradict that belief. However, because we asked the participants about their beliefs at the end of the experiment, it is possible that they had other beliefs while making the EOL judg-
ments, but changed their beliefs by the end of the experiment. Therefore, future studies may want to measure participants’ beliefs in the beginning of the experiment as well. It may also be possible to manipulate participants’ beliefs by informing them that some condition will make words easier to learn. Manipulation of beliefs has been successful in studies of JOLs (e.g., Mueller & Dunlosky, 2017), and it would likely work for EOL judgments as well.

Furthermore, in line with the inferential view of metacognition (e.g., Koriat, 1997), the metacognitive illusion we found showed that people may judge that a factor influence future learning even though it does not.
General Discussion

It is fascinating that people are able to foresee what will be easy or difficult to learn. Although this ability is likely an important and influential aspect of learning, only limited attention has been devoted to the issue. Indeed, Dunlosky and Metcalfe (2009, p. 91) noted that “sustained devotion to understanding EOL judgments has been rather minimal and little is known about them”. Almost a decade later, this description is still true. Hence, how people monitor their learning in the beginning of the learning process is still largely unknown. With the studies in this thesis, my colleagues and I have tried to alleviate some of the issues related to EOL judgments.

In the first sections, I will discuss the results from our three studies in relation to how accurate EOL judgments are. Thereafter, I will discuss what cues people use when they make EOL judgments, and how people use cues to make metacognitive judgments. I will mainly focus on processing fluency, which Study III added to the list of cues that influence EOL judgments. I will also discuss future directions of EOL judgment research and metacognition in general. Finally, I will end with a couple of concluding remarks.

The accuracy of EOL judgments

One of the major aims of this thesis was to investigate how accurate EOL judgments are and what moderates their accuracy. As outlined in the introduction, by comparing existing studies investigating EOL judgments, it is apparent that the accuracy of the judgments varies between studies. Some studies found that EOL judgments were moderately or highly accurate (Jönsson & Kerimi, 2011; Lippman & Kintz, 1968; Peynircioğlu et al., 2014; Underwood, 1966), while others found accuracy to be low (Jönsson & Kerimi, 2011; Kelmem et al., 2000; Leonesio & Nelson, 1990; Mazzoni et al., 1997). Furthermore, when one previous study directly compared several different judgments, the results showed EOL judgments were the least accurate, only slightly more accurate than what would be predicted by chance (Leonesio & Nelson, 1990).
How does the discriminability of a predictive cue affect EOL judgment accuracy?

We reasoned that some studies found EOL judgments to have low accuracy because they used materials in which all items were similar. That is, studies which found lower accuracy used materials consisting of similar items (e.g., Leonesio & Nelson, 1990), whereas studies that reported higher accuracy used materials with higher variability (e.g., Underwood, 1966). Therefore, in Study I, we tested whether one explanation for low accuracy may be that predictive cues did not vary sufficiently for people to discriminate between items. To test this, we had people judge, study, and be tested on two lists of word pairs. In one of the lists, the pairs varied highly in semantic relatedness, allowing participants to discriminate between items based on differences in semantic relatedness. In the other list, all word pairs had similar semantic relatedness, thus it would be harder to use this cue to discriminate between items. In other words, we manipulated the discriminability of the semantic relatedness cue. Indeed, our results clearly showed the accuracy of the EOL judgments was higher (mean $G = .74$) when semantic relatedness varied between items, than when it varied less (mean $G = .21$). This finding was also replicated in Study II. However, the accuracy was not as high as in Study I because we intentionally excluded the items with the lowest semantic relatedness in order to avoid ceiling effects.

Our results are fairly intuitive. When people are asked to make EOL judgments on a list of word pairs and notice the pairs vary in semantic relatedness, they most likely quickly identify this cue as a good way of discriminating between easy and difficult items. When the cue is semantic relatedness, which is related to learning (e.g., Dunlosky & Matvey, 2001), this is a good idea. Although people also use cues such as font size (Rhodes & Castel, 2008) and whether or not words are written in alternating upper- and lower-case letters or only upper-case letters (Study III), although these cues are not predictive of learning or memory. It is possible these cues are used because they are the only cues available. If items would vary in some other more predictive cue, it is possible these useless cues would be less influential (cf., Rhodes & Castel, 2008). I continue this discussion in a later section.

With near optimal circumstances, how accurate are EOL judgments?

Semantic relatedness between words in a word pair is a predictive cue for whether or not each pair will be learned (e.g., Dunlosky & Matvey, 2001). Therefore, in the high variability condition of Study I, and by using this cue, we attempted to create near optimal circumstances for our participants to be able to make accurate EOL judgments. Although the results showed the gamma correlations between judgments and memory performance on the test
was high (mean \( G = .74 \)), it was not perfect. This was most likely because properties of the pairs other than semantic relatedness influence whether or not an item would be learned. For example, during study, participants may have been concentrating more when studying some pairs compared to others. That is, idiosyncratic learning occurred which may have been hard to foresee when the EOL judgments were made. Because the EOL judgments were not reflecting these idiosyncratic influences on learning, some pairs that were judged as easy may not have been learned and some of the pairs judged as difficult may have been learned. Consequently, this resulted in lower than perfect correlations. However, EOL judgment accuracy was still high. When future studies investigate EOL judgments (and related judgments), researchers should consider what materials they are using, as this will have a large effect on judgment accuracy. This may also be important when researchers are investigating people’s study decisions. If people are unable to make accurate judgments (because predictive cues are not varying), they may have trouble making good study decisions.

**Does experimental methodology influence EOL judgment accuracy?**

In Study I, we also investigated the influence of three methodological aspects on the accuracy of EOL judgments. Namely, the judgment task, what criterion measure judgments were correlated with, and whether judgments were made with only one item visible, or all items at the same time. To investigate how EOL judgment accuracy was influenced by the judgment task, we asked participants to make judgments on a traditional difficulty scale ranging from easy to difficult, and a trials-to-learn scale on which they judged how many study trials they would need before learning the item. At the same time, we used two different criterion measures to assess how well participants learned each item. One criterion was binary and consisted of recall performance on the first test, while the other was a trials-to-learn criterion measuring how many study trials was needed to learn the items. As suggested by the notion of transfer-appropriate monitoring (e.g., Dunlosky & Nelson, 1992), we predicted that when the task and the criterion matched, as with the trials-to-learn judgment task and trials-to-learn criterion, accuracy would be higher. To be precise, accuracy may be higher if what people are asked to judge more closely matches the criterion measure. However, we found no such interaction. We also predicted that the trials-to-learn criterion may result in higher correlations than the binary criterion, because the binary criterion may reduce the correlation as a result of a restriction of range (Schwartz & Metcalfe, 1994). In contrast, we found that the binary criterion continuously produced higher correlations compared to the trials-to-learn criterion, independent of which EOL judgment task it was correlated with. However, the effect size was always small, indicating
that this difference is probably not influential on EOL judgment accuracy in general. One possible explanation for this unexpected difference in accuracy is that the binary criterion was closer in time to the EOL judgments, and thus was not affected by idiosyncratic learning that would occur after the first study trial. Meanwhile, the trials-to-learn criterion was affected by idiosyncratic learning in all study trials. According to the idea of test-potentiated learning (Arnold & McDermott, 2013), people will study items differently depending on how successful they were at recalling the items on a previous test. That is, if an item was not recalled on a test, people may focus more on learning that item compared to an item that was recalled. Participants were most likely not able to predict which items would be studied differently after the first test, and therefore their predictions better predicted the performance on the first test than the trials-to-learn criterion. If this explanation is correct, it suggests that EOL judgments are more predictive of early than later learning. Still, the difference between the binary and trials-to-learn criterion was small, and in everyday learning situations people can presumably correctly judge what will be easy or difficult to learn, even when they plan to study for a long time. At least when the material varies in some predictive cue.

We also investigated whether there was any difference in accuracy when judgments were made with items presented simultaneously or sequentially. We expected there to be an advantage in accuracy when items were presented simultaneously as this would allow participants to compare all items without having to remember the other items. However, we found no such difference. One possible explanation for this may be that people continuously update an average item difficulty, and relate their judgments to this average. Thus, they did not need to see all items to judge how difficult an item was compared to other items. However, it is also possible that it was sufficient for people to compare each item with the previous items that came to mind, or an average based on previous pre-experimental learning.

Overall, the results from Study I indicate that EOL judgments may be surprisingly robust to changes in the circumstances in which the judgments are made. It is possible that slight variations to EOL judgment tasks and presentation methods do not result in measurable differences in EOL judgments, suggesting that these judgments are consistently applied by participants. Nonetheless, we compared somewhat small changes in experimental method and large changes may still influence results. In conclusion, our results show that researchers can most likely compare the accuracy of EOL judgments between studies with different methodology, as long as they use similar materials.

Can EOL judgments be as accurate as JOLs?

In Nelson and Narens’ (1990) theoretical framework, they describe how people continuously monitor their learning and memory throughout the learning process. In the beginning of learning people may monitor how easy or difficult
material is to learn (i.e., EOL judgments), and later they may monitor how well they have learned the material (i.e., JOLs). Although differently timed in the learning process, EOL judgments and JOLs are similar and likely often rely on similar cues. That is, based on available cues, predictions are made about future memory states, before study and after study. However, because JOLs are made after study, additional cues are available, which may help JOLs be more accurate than EOL judgments. Although Study I, Study II, and older studies (e.g., Jönsson & Kerimi, 2011; Underwood, 1966) show EOL judgments can be accurate, it is possible that JOLs will always be more accurate than EOL judgments because they have access to idiosyncratic cues signaling how well items have been learned during study. That is, JOLs can, as EOL judgments, rely on information from item characteristics, but also on cues signaling whether one item has been better learned than another. For example, JOLs could rely on whether one item feels more familiar than another or is easier to retrieve from memory (e.g., Metcalfe & Finn, 2008b).

As mentioned, when the accuracy of EOL judgments and cue-target JOLs (made with both cue and target words visible) were directly compared in the past, cue-target JOLs were more accurate (Leonesio & Nelson, 1990). However, the material used in that study did not vary in any predictive characteristic. Because no predictive characteristics were available, the cue-target JOLs were likely more accurate because they could rely on cues signaling how well the items had been learned during study. To test whether a predictive cue could allow EOL judgments to be as accurate as cue-target JOLs (or make the difference smaller), we had participants make both EOL judgments and cue-target JOLs on two lists of word pairs (Study II). As in Study I, the word pairs in one list varied highly in semantic relatedness and in the other it varied less. Indeed, as predicted, there was no reliable difference in accuracy when they varied highly in semantic relatedness. Meanwhile, our results showed that cue-target JOLs were more accurate when the material had low variability in semantic relatedness. That is, in the low variability condition, we replicated Leonesio and Nelson’s (1990) result that EOL judgments are less accurate than JOLs. Meanwhile, in the high variability condition, we showed that EOL judgments can be as accurate as cue-target JOLs. This shows Leonesio and Nelson’s (1990) finding has limited generalizability. If EOL judgments can rely on salient and predictive cues that can be used to discriminate between items, they can be as accurate as cue-target JOLs. Furthermore, correlations between judgments and the pre-standardized semantic relatedness of the word pairs, indicate that the cue-target JOLs mainly relied on semantic relatedness as well. That is, in the high variability condition, EOL judgments and cue-target JOLs were strikingly similar in both accuracy and how much they relied on semantic relatedness. This could indicate that cue-target JOLs did not rely on idiosyncratic cues.

Another possibility is that cue-target JOLs do not rely on semantic relatedness as much as EOL judgments, but the idiosyncratic cues were highly
correlated with the semantic relatedness. That is, because the idiosyncratic cues and the semantic relatedness signaled the same items as being better learned, the correlation between cue-target JOLs and semantic relatedness was equally high for the EOL judgments. However, even though semantic relatedness is predictive of learning and memory, it is not perfect. If cue-target JOLs relied on both semantic relatedness and idiosyncratic cues, they could probably have been more accurate than EOL judgments. Although most idiosyncratic cues are available after study, it is also important to note that some idiosyncratic cues may be available when people make EOL judgments. For instance, because some learning may occur while items are judged, people may be monitoring how fluently the items are encoded to memory (see, Hertzog et al., 2003). For example, they may be monitoring how easy it is to create an association between the words in the pairs, or how easy it is to create a mental picture with the two words together. However, most likely, these cues only have small influence on EOL judgments, or are not very predictive of learning, as we would otherwise have found higher accuracy in the low variability conditions in Study I and II. Because both EOL judgments and JOLs presumably affect how people regulate their study (Nelson & Narens, 1990), Study I and II indicate that people may not need to study a material before they plan their study time. Instead, people can probably make good plans in the beginning of learning as well, at least if the material varies in difficulty. However, to what degree planning one’s study time in the beginning of the learning process benefits learning has never been tested.

Despite the similarities in accuracy between the EOL judgments and cue-target JOLs, there is one large advantage of JOLs over EOL judgments. After study, people can test their memory. In Study II, we compared EOL judgments with JOLs made on only the cue-word (e.g., sun – ?) of the cue-target word pair (e.g., sun – warm) and this made the cue-only JOLs highly accurate (median $G = .92$) and more accurate than the EOL judgments, independent of material composition. It is well established in earlier research that making JOLs in combination with a retrieval attempt makes them highly accurate (e.g., Nelson & Dunlosky, 1991), and our cue-only JOL condition allowed just that. In other words, in a delayed cue-only JOL condition, people may test their memory to know whether or not an item is learned, and the outcome of such diagnostic retrieval is a predictive cue for whether or not an item will be recalled later. Hence, when people can test their memory, JOLs will, in most cases, be more accurate than EOL judgments. However, it is likely possible to create a material which varies sufficiently in a predictive characteristic to allow EOL judgments to be as accurate. In addition, it may even be possible to make EOL judgments more accurate than JOLs by creating a metacognitive illusion which only affects JOLs. Still, in many everyday learning situations, JOLs will likely be more accurate, because they can rely on additional predictive cues.
The underlying basis of EOL judgments

One of the main aims of this thesis was to investigate what cues people use when they make EOL judgments and how they use cues to make metacognitive judgments. In this section I will discuss research questions related to this aim, and how the results of my studies help answer parts of the questions.

Do cues compete for influence?

Besides showing that EOL judgments may be as accurate as cue-target JOLs, Study II indicates that cues compete for influence over judgments. With the cue competition hypothesis, we proposed that one cue may alter the influence of another cue. The results of Study II support this idea, as there was no reliable difference in accuracy between EOL judgments and cue-target JOLs when the judged word pairs varied in semantic relatedness. Meanwhile, when semantic relatedness did not vary, cue-target JOLs were more accurate than the EOL judgments. This indicates that participants relied on cues other than semantic relatedness when they made cue-target JOLs on the low variability list. Most likely the other cues were idiosyncratic cues signaling how well each item had been learned during study. Because cue-target JOL accuracy was not higher than EOL judgment accuracy when semantic relatedness varied, this indicates that JOLs relied on idiosyncratic cues less when semantic relatedness varied. If participants had used both semantic relatedness and idiosyncratic cues when they made cue-target JOLs on the high variability list, they should have been more accurate than EOL judgments. However, it is possible the semantic relatedness was sufficiently salient and predictive that participants more or less ignored idiosyncratic cues. This explanation for our results was also supported by the correlations between judgments and pre-standardized semantic relatedness, as the correlation did not differ for EOL judgments and cue-target JOLs. This is of theoretical importance because it suggests, in line with the cue competition hypothesis, that cues compete when people monitor learning and memory. It is possible that, if people identify a salient and predictive cue, they do not search for other cues that may help them to discriminate between items. Another possibility is that many cues are identified but people primarily use the most salient and predictive cue, limiting the influence of the other cues.

The results of Study II indicate that cue-target JOLs did not, or only to a lower extent, rely on idiosyncratic cues when the judged word pairs varied in semantic relatedness. Together with earlier studies indicating that a cue has smaller influence when another cue is available (e.g., Rhodes & Castel, 2008), the results of Study II suggest cues compete for influence over metacognitive judgments. One reason for this could have been that the idiosyncratic cues seemed less predictive when semantic relatedness varied. In the judgment and decision-making literature, there is evidence that the presence of one cue can
alter how predictive another cue seems. When Price and Yates (1993) investigated the phenomenon of judgmental overshadowing, they found that a moderately predictive cue was judged to be less predictive when a second cue had high predictive power, whereas, when the second cue had low predictive power, the first cue was judged to be more predictive. Thus, how predictive a cue seems can be altered by other cues. Perhaps this is why cues compete for influence over metacognitive judgments. Semantic relatedness is predictive of learning and memory (e.g., Dunlosky & Matvey, 2001), and perhaps idiosyncratic cues seemed less predictive when semantic relatedness varied. Consequently, participants may have relied less on idiosyncratic cues because they seemed less predictive.

However, it is possible cues used for judgments of one’s own learning and memory are not affected by other cues in the same way as cues used for other types of judgments. When people judge their learning and memory, many cues participants rely upon are likely used because the participants have learned throughout life that these cues are predictive of learning and memory. In Price and Yates’ (1993) experiments, participants learned how predictive the cues were during the experiment. If people have learned how predictive cues are throughout life, then perhaps their seemed predictive power is not altered by other cues. Instead, there may be other explanations for why cues compete for influence over metacognitive judgments related to memory and learning. Perhaps a cue becomes more influential because it is relatively more salient or predictive compared to other available cues, not because one cue makes other cues seem less predictive. Why cues compete should be investigated in future studies.

The results of Study II are in line with what is predicted by the cue competition hypothesis. However, there is an alternative explanation which could potentially explain our results. It is possible participants used both cues when they made cue-target JOLs, but both cues signaled that the same items would be more or less likely to be remembered. For example, if items with high semantic relatedness were also the items that were best learned after study, there would be no additional benefit to accuracy of relying on both cues. One way to investigate whether both idiosyncratic cues and semantic relatedness are used at the same time could be to lower the predictive power of semantic relatedness but keep the high variation in semantic relatedness. This may be achieved by having participants study items for longer before they make cue-target JOLs. If each item would have been studied for a longer time, the predictive power of semantic relatedness would likely be lower, as idiosyncratic learning would have a larger influence on which items are remembered on the test. Consequently, with more study time, cue-target JOLs could have higher accuracy than EOL judgments because the idiosyncratic cues may have relatively higher predictive power compared to semantic relatedness. However, if EOL judgments and cue-target JOLs would still not differ in accuracy, this
would support the idea that participants primarily favored using semantic relatedness over idiosyncratic cues, perhaps because semantic relatedness is more salient. In Study II, EOL judgments and JOLs were only separated by one 5-second study trial per item. Hence, it is possible cue-target JOL accuracy would be higher than EOL judgment accuracy if each item had been studied for a longer time before JOLs were made.

In Study II, participants also made cue-only JOLs, and in this condition accuracy was higher than both EOL judgments and cue-target JOLs, independent of how much semantic relatedness varied. Correlations between judgments and semantic relatedness also showed people relied less on semantic relatedness than for the other judgment types. That is, when participants tried to recall the target word during the judgment, they primarily relied on whether or not they could access the word from memory and the judgments had a low correlation with semantic relatedness. However, this only partly supports the cue competition hypothesis, as participants only had access to the semantic relatedness of the words when they were able to correctly recall the target word from memory during the judgments. Further investigating how multiple cues compete to influence metacognitive judgments will be important for developing theories for how people use cues when they make metacognitive judgments.

Understanding how people use multiple cues could also have practical use in everyday learning situations. In many learning situations, learners likely have simultaneous access to multiple cues. If cues do compete (which Study II indicates), it may be important for students to know which cues they should rely upon. Perhaps it would be advisable for people to ignore their intuitive feeling for which cues to use. For example, in Study III, relying on processing fluency as a cue resulted in inaccurate EOL judgments. This said, it might be difficult to change what cues people rely upon, by only informing them to ignore a cue. When Rhodes and Castel (2008, experiment 4) explicitly informed their participants that font size was unrelated to future memory performance, font size still had a small impact on JOLs. That is, their instructions did not prevent participants from using font size as a cue. However, in their experiment, they only manipulated font size. It is possible that if they had manipulated two cues at the same time, participants might have followed the instructions and ignored the font size, in favor for the other cue.

One limitation with Study II was that we only had control over one cue. Specifically, we had control over semantic relatedness, but the additional idiosyncratic cues were made available while studying, and were most likely also correlated with semantic relatedness. Future studies should more directly investigate how multiple cues may compete with each other. One way of manipulating how a cue is weighted could be to manipulate participants’ beliefs about the cue. For example, this could be achieved by informing participants that one cue has a larger impact on memory or learning at the beginning of an
experiment. A similar method has been successfully used by Mueller and Dunlosky (2017) to influence how people have used a single cue. In their study, they informed participants that the brain has an easier time processing either the color blue or green. Participants subsequently studied and made JOLs for items presented in green and blue font color, and their judgments were affected by the instructions. Accordingly, informing participants that one cue (e.g., font size) has a larger impact on learning than another cue (e.g., font color) could possibly influence how participants weigh the cues when they make metacognitive judgments.

Furthermore, Undorf et al. (2018) recently showed that people are able to integrate up to four cues when they make JOLs. However, even though four experimentally manipulated cues affected the judgments, it is possible participants did not use all four of them directly. Instead it could mean participants are relying on one cue (or more) with which the other cues are correlated. Processing fluency may be a good candidate for such a cue. That is, it is possible that many cues could influence processing fluency and people’s judgments would be correlated with each cue separately even though they primarily relied on processing fluency. Investigating this possibility could be an important route for further developing theories of how people use multiple cues to make metacognitive judgments.

In conclusion, by showing that cue-target JOLs were more accurate than EOL judgments when semantic relatedness did not vary, but equally accurate when it did vary, the results support the cue competition hypothesis. Consequently, Study II indicates that that cue-target JOLs relied on idiosyncratic cues when semantic relatedness did not vary, but they relied less on these cues when semantic relatedness varied. At this point, there are no theories on how cues are weighted when people judge their own learning and memory, and these results highlight the importance of developing such theories to understand how people use multiple cues when they make metacognitive judgments.

Is processing fluency used as a cue for EOL judgments?

Research on metacognitive monitoring supports the view that monitoring is inferential (Koriat & Lieblich, 1977; Metcalfe, 2000; Metcalfe et al., 1993; Reder & Ritter, 1992; Schwartz & Smith, 1997). When people monitor their learning and memory they rely on cues available in the current learning situations to predict how easy or difficult an item will be to learn or remember. In this section I will discuss which cues affect EOL judgments, with a focus on processing fluency.

Research on EOL judgments show that people’s judgments are mainly based on characteristics of the judged material. That is, EOL judgments are influenced by cues such as semantic relatedness between words in a word pair (Study I and II), word length and concreteness of the word (Jönsson &
Lindström, 2010), or how often a word is used in a language (Begg et al., 1989; Jönsson & Lindström, 2010), the meaningfulness of an item (Lippman & Kintz, 1968), or how typical it is of its category (Mazzoni et al., 1997). However, people may also use information about how the material is presented (Peynircioğlu et al., 2014) and how other people have performed when learning the same material (de Carvalho Filho & Yuzawa, 2001). Furthermore, EOL judgments are also affected by how proficient people are in the type of material they are learning (Löffler et al., 2016).

Although these cues have been identified in laboratory studies, most of them likely also affect people’s judgments of their learning in everyday learning situations. However, in everyday learning situations, people often learn more complex materials than word pairs or single words. In more complex materials, other cues are also likely to be important. For example, if a student quickly looks through a textbook to assess which chapters will be more or less difficult to learn, the student may rely on chapter length, how complex the writing seems, or perhaps how many pictures there are in each chapter. Nevertheless, people probably use these cues in the same way as cues in less complex materials. Hence, findings of how cues are used in laboratory studies are likely to generalize to everyday learning situations.

In Study III, we added another cue to the list of cues that people use when they make EOL judgments, namely processing fluency. This follows from earlier work (Begg et al., 1989). In the 1980s, Begg et al. (1989) varied word frequency (how often a word was used in the English language) in a list of words. They did this because they suspected word frequency would affect how easily or fast each word would be to process. Their results indicated that word frequency indeed affected EOL judgments. However, because they did not directly manipulate processing fluency, their results could also have been the result of some other unknown aspect related to word frequency. Nevertheless, as Begg and his colleagues suggested, Study III confirmed that EOL judgments are influenced by processing fluency. We manipulated processing fluency by randomly assigning words to be presented in a condition associated with low processing fluency (alternating condition) or high processing fluency (constant condition). In the alternating condition, words were presented with alternating upper- and lower-case letters (e.g., bUcKeT). In the constant condition, words were instead presented in only upper-case letters (e.g., BUCKET). Consistent with the view that processing fluency affects EOL judgments, participants judged words in the constant condition as easier to learn than in the alternating condition. They did this even though the manipulation had no effect on memory performance. This manipulation has been used in earlier research to manipulate processing fluency (e.g., Rhodes & Castel, 2008). Nevertheless, to control that our manipulation actually affected processing fluency, we compared how much time participants used to complete the EOL judgments across the two conditions. The results showed participants used more time to judge words presented in the alternating condition than in
the constant condition. Hence, the difference in judgment time indicates that our manipulation most likely affected processing fluency. However, it is unclear exactly what type of fluency was affected (for a review see, Alter & Oppenheimer, 2009). For example, the manipulation may not have only affected perceptual fluency but could also have affected reading speed. In addition, because participants were exposed to the words while they judged them, some learning likely occurred, and they may have monitored how easily the words were to encode to memory (Hertzog et al., 2003).

According to the inferential view of metacognition, memory and learning are dissociable from monitoring. In so doing, monitoring can be influenced by cues that do not affect learning. Several studies have demonstrated metacognitive illusions in JOLs (e.g., Rhodes & Castel, 2008). In Study III, we showed that illusions can occur for EOL judgments as well. Participants in our study continuously judged words in the alternating condition as being less likely to be learned than words in the constant condition, even though our manipulation had no effect on whether or not the words would be recalled later. One possible explanation as to why the participants relied on the processing fluency is that we used a homogenous material. Specifically, all of the words in the study were six-letter concrete nouns with similar word frequency. It is likely that processing fluency was the most salient cue that could be used to discriminate between the words, and therefore the participants used it. It is possible that our participants would have ignored the difference in processing fluency if a more salient and predictive cue would have been available.

How do people’s beliefs about a cue impact how it is used?
In Flavell’s (1979) influential article, he argued that metacognitive knowledge (i.e., beliefs) affect how cues are interpreted to affect our cognitive enterprises. For example, if a student believes that it is easier to learn the association between two semantically related words than two unrelated words, their judgments will follow their belief. According to the analytic processing theory (Mueller & Dunlosky, 2017), monitoring is affected by cues because people have beliefs about the cues. Specifically, when people are asked to make, for example, a JOL, they search for available cues and then respond according to their beliefs about the found cues. This theory has support in the literature. Recent studies have demonstrated that people’s beliefs mediate how font size (Mueller et al., 2014) and semantic relatedness (Mueller et al., 2013) influence monitoring. It has also been shown that it is possible to change how the color of words impacts JOLs, by manipulating people’s beliefs about how the brain processes color (Mueller & Dunlosky, 2017). However, other studies have shown that beliefs sometimes cannot explain the full impact of a cue on monitoring. In one study, participants judged words presented in a high volume as being more likely to be remembered than words presented in a lower volume, even though this was contrary to participants’ beliefs (Frank & Kuhlmann,
If a cue can influence monitoring without being based on beliefs, then processing fluency is a good candidate for how this could work.

To investigate the combined effect of processing fluency and beliefs, we investigated how processing fluency and beliefs interacted in influencing EOL judgments in Study III. Our results showed that beliefs do indeed impact EOL judgments, but they could not explain all of the influence of our manipulation (similar to Frank & Kuhlmann, 2017). Specifically, many participants judged words in the constant condition (high fluency) as more likely to be learned than words in the alternating condition (low fluency), even when it contradicted their beliefs. Still, beliefs and processing fluency interacted. The results showed that the size of the difference between words presented in the alternating and constant condition was affected by beliefs. More specifically, how much higher EOL judgments were for words in the constant condition, compared to the alternating condition, depended on beliefs. The size of the difference was largest when people believed words in the constant condition were more likely to be learned, followed by believing there was no difference, and finally believing words in the alternating condition were more likely to be learned. Thus, our results indicate that beliefs moderate the effect of processing fluency, but not that it mediated the effect.

It is possible that beliefs mediate the influence of some cues on monitoring, but Study III showed that this is not always the case. Based on our results, it is difficult to say why some experiments have found that beliefs mediate the influence of a cue (e.g., Mueller et al., 2013), while ours did not. One possibility is that EOL judgments are more sensitive to processing fluency than other metacognitive judgments. One way of investigating this would be to have people make both EOL judgments and JOLs on the same material as was used in Study III and investigate whether the two types of judgments are affected differently. However, given similarities between EOL judgments and JOLs, it is difficult to say why people would use beliefs differently for the two types of judgments.

Another possibility for why we did not find that beliefs mediated the effect is methodological in nature. In Study III, participants reported their beliefs at the end of the experiment. It is possible that some people changed their beliefs throughout the experiment. That is, people may have believed that words presented in the constant condition would be easier to learn in the beginning of the experiment, but, sometime during the experiment, some of the participants changed their beliefs. However, our results match what would be expected from the three belief groups. Namely, we found that the difference between the alternating and constant condition was largest for the group who believed the constant condition made the words more likely to be learned, followed by the group who believed there was no difference, and smallest for the group that believed the alternating condition made the words more likely to be learned. Hence, although some participants may have changed their beliefs throughout the experiment, this probably does not no explain the results found...
in Study III. Nevertheless, one simple way of testing whether this methodological aspect could explain our results would be to ask participants about their beliefs both in the beginning and at the end of the experiment. Another way could be to use a method similar to experiments that have used so-called pre-study JOLs (Mueller et al., 2013). Specifically, participants would be told that an upcoming word would be presented either in the constant or alternating condition, and then make an EOL judgment based on this information alone without viewing the item. If this procedure resulted in the same pattern of results as found in Study III, we might have underestimated the effects of beliefs on EOL judgments.

Theoretically, the results of Study III, together with prior research, imply that beliefs guide the influence cues such as semantic relatedness have on monitoring, but that processing fluency has an effect on monitoring even when it beliefs run counter to the direction of processing fluency. Perhaps low processing fluency is so strongly associated with difficulty that it circumvents beliefs. Consequently, it is possible that many cues, such as semantic relatedness or word frequency, can both influence metacognitive judgments through beliefs and through how the cue affects processing fluency. For example, if common words are easier to process than uncommon words, perhaps judgments are automatically influenced by this difference in processing fluency. At the same time, people may believe that common words are easier to learn and then apply this belief when they make EOL judgment. In this way, word frequency could influence judgments through both beliefs and processing fluency at the same time. It may be possible to test whether this is the case by separately manipulating processing fluency and, for example, word frequency to see how beliefs and processing fluency affect monitoring separately. However, separating processing fluency from beliefs will be problematic, as people can also have beliefs about processing fluency (Mueller & Dunlosky, 2017). Hence, if a manipulation affects processing fluency, it could impact judgments because people believe processing fluency has an effect.

Nevertheless, given that both beliefs and processing fluency can influence metacognitive judgments, one interesting possibility is that judgments are made in two stages. Initially, properties of a learning situation and material, such as word frequency, affect monitoring through processing fluency. People then actively search for properties that can be used as cues and update their judgments based on how they believe the properties affect learning and memory (cf., Mueller & Dunlosky, 2017). There is support for a two-stage model of how judgments are formed when the combined effect of familiarity and memory retrieval have been investigated (Metcalf & Finn, 2008b). Specifically, when people experience a strong feeling of familiarity (likely based on processing fluency), they try to retrieve the missing item from memory, resulting in a slow judgment. However, if familiarity is low, they make a fast judgment, based on the familiarity. Similar results have been found when people make FOK judgments (Koriat & Levy-Sadot, 2001). Hence, monitoring
went through two stages, first familiarity then retrieval. Investigating the possibility that all metacognitive judgments, not just judgments made in combination with memory retrieval, go through two stages, may lead to great advances in the theoretical understanding of metacognitive monitoring.

**Further directions**

In the above sections, I discussed some future direction of research on metacognitive monitoring. In the following section I continue to discuss future directions that could advance the theoretical understanding about metacognition and also help the development of guidelines for how people should judge their learning and memory to make their learning more efficient.

**Ease of learning judgments and study regulation**

Both Study I and II showed that EOL judgments are accurate under some circumstances. There are studies that show EOL judgments are correlated with how much study time is distributed to different items (e.g., Nelson & Leonesio, 1988; Son & Metcalfe, 2000). However, there is no research on whether students who plan their study time in the beginning of learning learn more. This is important to investigate because it could show whether planning one’s study time in the beginning of learning is advisable for students in an everyday learning situation. In addition, there is no research showing that EOL judgments are causally related to study regulation in the beginning of learning. However, most likely people rely on the metacognitive monitoring in the beginning of learning when they plan how to spend their study time. Studies on JOLs have shown they are causally related to study regulation (Metcalfe & Finn, 2008a), and given the similarities with JOLs, EOL judgments should be as well. Future research could make use of metacognitive illusions to show how EOL judgments are related to study regulation, similar to research on JOLs (e.g., Metcalfe & Finn, 2008a). Hence, there are several avenues of future research when it comes to showing the relationship between monitoring in the beginning of learning and how it affects study regulation and learning.

Although both Study I and Study II showed that EOL judgments can be accurate, we also showed that EOL judgments are sometimes inaccurate (Study I-III). Because of this, it may be a good idea for teachers and authors of textbooks to guide students in how they should best allocate their study time. Perhaps it could be possible to create recommendations showing how long the average student usually need to study a chapter to learn its content. That way, students could rely on these recommendations when they allocate their study time, rather than their potentially inaccurate monitoring. Because EOL judgments are influenced by information about other people’s memory performance on the same material (de Carvalho Filho & Yuzawa, 2001), it is...
likely that peoples monitoring and study regulation would be influenced by such recommendations. Nevertheless, this should be tested to show such instructions would have a benefit to learning before it is recommended to teachers and authors of textbooks.

Another interesting line of research would be to investigate whether people know when their monitoring will be accurate or inaccurate. That is, are people aware of when they will be able to make accurate predictions, and do they know when a material includes predictive cues? To the best of my knowledge, this has only been tested for JOLs, and the research indicates that people prefer using less accurate cue-target JOLs instead of the more accurate cue-only JOLs (Jönsson & Kerimi, 2011; Todorov, Kornell, Larsson Sundqvist, & Jönsson, 2013). This indicates that people are unaware of what conditions will allow them to accurately make JOLs, and this may be true for EOL judgments as well.

If people do know when they can and cannot make accurate judgments, they may know when they can plan ahead in the beginning of learning and when they should study the material before planning their study time. However, if they do not know this, they should perhaps be advised to study the material for longer, before planning how to spend their study time, just to be on the safe side. By this, I mean that future studies may want to investigate if people can accurately predict when their EOL judgments will be accurate.

Individual differences in monitoring ability

There has been little research on whether people vary in monitoring ability and whether they have a general monitoring ability at all (but see, Kelemen et al., 2000). That is, it is still unclear whether some individuals’ monitoring is continuously more accurate than other individuals. There are studies showing individual differences in metacognitive knowledge (e.g., Efklides & Vlachopoulos, 2012; Schraw & Dennison, 1994), and differences in judgment accuracy depending on expertise in the judged type of material (Löfler et al., 2016). However, to the best of my knowledge, only one study has directly tested whether some individuals have a better monitoring ability than others by using a test re-test method (Kelemen et al., 2000). Kelemen et al. (2000) first had participants complete several types of metacognitive judgments (including EOL judgments and JOLs). The same participants then returned a week later and completed the same tasks a second time (but on a new material). The results showed that the participants’ monitoring accuracy was not stable across tasks or the two sessions. Hence, there was no support for a general monitoring ability. However, some of the judgment types were consistently more accurate. This led the authors to argue that monitoring may be state-dependent rather than depend on a metacognitive ability. In other words, people will not be consistently more accurate than others, but people will, in general, be more accurate when they make, for example, delayed cue-only
JOLs, compared to cue-target JOLs (e.g., Nelson & Dunlosky, 1991; Study II). However, a later study showed that individuals do differ in which cues they use to make JOLs. The study showed that people could integrate up to four cues when they made JOLs (Undorf et al., 2018). Interestingly, the study also showed that some participants used fewer than four cues and some participants used none of the manipulated cues. This suggests that people differ in how they integrate multiple cues. People relying on a different number of cues could perhaps be both a result of which strategy they use, but also how many cues they have the capacity to utilize at the same time. If some individuals are better at using multiple cues, then this could potentially show that some people are consistently more accurate than others. Undorf et al., (2018) investigated this and found positive correlations between number of used cues and accuracy in one of their four experiments, but not in the other three experiments. However, their experiments were not primarily designed to investigate this question, and all of the manipulated cues were not predictive of memory performance. Hence, more research is needed, and it is possible that if Kelemen et al. (2000) would have used a material varying in multiple predictive cues, they may have found evidence for a general monitoring ability.

Related to this idea, it is possible that Kelemen et al. (2000) did not find support for a general monitoring ability because they used a material which did not vary in any predictive cues. In their study, they specifically selected items to be similar in difficulty. Consequently, they also found fairly low correlations between judgments and memory performance (except when participants made delayed cue-only JOLs). As we showed in both Study I and Study II, people’s judgment accuracy is highly dependent on how much the available cues vary. It is possible that if the material lacks predictive cues that vary, then individuals with good monitoring ability do not have the chance to be better at using the cues than people with worse monitoring ability. If the material instead had included a highly predictive cue (e.g., Study I and II), then participants with better monitoring ability would perhaps have been continuously more accurate than others over both tests and sessions. Although monitoring is likely state-dependent, as Kelemen et al. (2000) argued, there are likely individual differences that affect individuals’ ability to monitor their memory and learning. However, it is yet unclear what individual differences this could be, but possible candidates are what strategy people use, and perhaps a capacity to integrate and correctly weight multiple cues (cf., Undorf et al., 2018). Finding a reliable method to identify differences in monitoring ability could lead to understanding of whether people differ, and perhaps also for developing methods to help people improving their monitoring ability.
Concluding remarks

In the three studies included in this thesis, I showed that EOL judgments can be accurate (Study I and II), and in some conditions, they can even be as accurate as JOLs (Study II). Furthermore, the main reason some older studies have found that EOL judgments are inaccurate (e.g., Kelemen et al., 2000) and the least accurate among several judgment types (Leonesio & Nelson, 1990), seems to be that these studies used materials in which items did not vary with respect to predictive characteristics that EOL judgments could rely upon. In addition, Study II also indicates that cues may compete; that is, one cue becomes less influential when another salient and predictive cue is available. Finally, Study III also showed that processing fluency may be an important cue for EOL judgments. However, the third study also showed that people’s beliefs about the manipulation will moderate the impact of processing fluency.

There is great potential for applying metacognitive research into everyday learning situations. Several findings from the literature have been translated into learning advice, for instance the effect that people should test their memory to accurately monitor what they have learned (e.g., Nelson & Dunlosky, 1991). Because research on EOL judgments is still limited, there are many unexplored avenues that could potentially be applied to everyday learning. Perhaps there are similar ways of improving monitoring accuracy in the beginning of the learning process as there is for later in the learning process. Additional research is needed about metacognitive monitoring and control at the beginning of the learning process in order to potentially show these effects. I look forward to being part of the research in this interesting line of research.
Acknowledgements

First and foremost, I’m deeply thankful to my main supervisor Fredrik Jöns-
son. Without your guidance, every step throughout this process would have
been much more difficult. I’m especially grateful for all the times you listen
to my endless ideas (of varying quality) and nudged me in the right direction.
Thank you for inviting me to work with you, it has been a wonderful experi-
ence.

To my other supervisors, thank you for all your help and valuable advice.
Veit Kubik, I don’t know where you find your energy, but it is inspiring.
You’ve been a great colleague ever since I started as a research assistant and
I’ve really liked sharing ideas and working on our projects together. Timo
Mäntylä, thank you for your support and all your great advice on how to im-
prove what I’ve written. Last but not least, Bennett Schwartz. It has been won-
derful working with you. Coming to visit you in Florida was one of the best
decisions I made as a student. You really made me feel welcome and I had a
great time. I look forward to continuing working together.

A big thank you to all my dear colleagues and friends at the department.
The last years have really been great, both outside and at work. I’ve specially
enjoyed all the lunch conversations, bathing in Brunsviken, and the hours of
board gaming. I’m happy to say I now consider many of you not only as won-
derful colleagues but as friends.

I would also like to thank all my friends and family. You are what make
every day meaningful. I’m so happy to have each and every one of you in my
life. All of you have helped me even if you don’t know it. Some extra thanks
to Johan, for showing me that it is possible to become great at anything if you
just put your mind to it. Also, Christina, for helping me believe I could do
research. Mom and dad, thank you for all your love and support, and the op-
portunities you’ve given me. Finally, Malin, regardless of my mood through-
out the day, I only needed to speak with you for a few moments to be as happy
as I can be. You have made this experience so much easier. I’m ever grateful
for your support and encouragement.
References


