On the Link between Fact Learning and General Cognitive Ability

Florian Sense (f.sense@rug.nl)
Department of Experimental Psychology & Department of Psychometrics and Statistics
Groningen, The Netherlands

Rob R. Meijer (r.r.meijer@rug.nl)
Department of Psychometrics and Statistics
Groningen, The Netherlands

Hedderik van Rijn (d.h.van.rijn@rug.nl)
Department of Experimental Psychology & Department of Psychometrics and Statistics
Groningen, The Netherlands

Abstract

Adaptive fact learning systems have been developed to make optimal use of testing and spacing effects by taking into account individual differences in learning efficiency. Measures derived from these systems, capturing the individual differences, predict later performance in similar *and* different fact learning tasks. Additionally, there is a rich body of literature showing that individual differences in general cognitive ability or working memory capacity can predict scores on achievement tests. If these measures also influence fact learning, incorporating them might further enhance adaptive systems. However, here we provide evidence that performance during fact learning is neither related to working memory capacity nor general cognitive ability. This means that the individual differences captured by our adaptive learning system encapsulate characteristics of learners that are independent of their general cognitive ability. Consequently, adaptive learning methods should focus primarily on memory-related processes.

Keywords: learning; memory; working memory capacity; general cognitive ability; fluid intelligence; individual differences; computational modeling

Introduction

Research has shown that standardized measures of acquired knowledge are among the best predictors of success in life (e.g., Kuncel & Hezlett, 2007, 2010). With an ever-growing body of knowledge, life-long learning has become a reality for many and acquiring new knowledge efficiently is paramount. Computerized learning systems are developed to streamline the acquisition of new knowledge and the most successful ones stand out because they adapt to the individual characteristics of the learner. What is not clear, however, is whether only individual differences in memory-related processes are relevant to optimize the adaptation or whether individual differences in general cognitive ability should be taken into account as well. Here, we present data that suggest that individual differences in general cognitive ability are not required to optimize the fact-learning process using the model developed in our lab.

Few findings in psychology are as reliably reproduced as the spacing effect (Donovan & Radosevich, 1999), the finding that learning yields better long-term results if repetitions are spaced over time rather than crammed together. The spacing effect holds over various time scales (e.g., Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008), indicating that it reflects a fundamental property of the human memory system. Therefore, the spacing effect should be exploited when we want to optimize fact learning (Dempster, 1988). Pavlik and Anderson (2003) extended ACT-R's declarative memory module to account for spacing effects and subsequently made first steps in using the decay-based model of human memory to enhance the learning of facts (Pavlik & Anderson, 2005, 2008). Their model uses trial-by-trial information gathered *during* learning to devise personalized learning schedules on the fly.

Van Rijn and colleagues (2009) refined the model further but the underlying principles are still the same: each item is assigned an activation value when it is first presented to the learner. The activation decays over time and the model strives to schedule a presentation of the item before the activation is too low, which would prevent a successful retrieval from memory. Throughout the learning session, the estimation of each item's activation is continuously fine-tuned based on the learner's response times to quiz-item. Specifically, on each trial, the model's predicted activation can be converted to an *expected* response time, which is compared to the *observed* response time. The discrepancy between the two is used to update the estimated decay rate of the activation. This way, one decay-related parameter is estimated for each item for each learner. By averaging across the item-specific parameters, we can compute a value that indicates how quickly, on average, a learner forgets the items in a particular set. We will refer to this value as the *rate of forgetting* (for detailed descriptions of the model see Sense, Behrens, Meijer, & Van Rijn, 2016; Van Rijn et al., 2009).

Nijboer (2011) has shown that the *rate of forgetting* estimated during learning of one set of items is very strongly related to subsequent test performance for the same set of items and the data presented here confirm this strong relationship. More recently, we showed that the *rate of forgetting* is stable over time within one type of material, and relatively stable over materials, which indicates that the *rate of forgetting* can be estimated reliably (Sense et al., 2016). Taken together, this suggests that someone's *rate of forgetting* is a useful and reliable measure of individual

differences. What is not clear, however, is what exactly the *rate of forgetting* encapsulates. Given the parameter's roots in ACT-R's declarative memory module (Anderson, Bothell, Lebiere, & Matessa, 1998; Anderson, 2007), one would expect that it captures memory-related processes. However, for a participant to do well in a laboratory simulation of a fact-learning session, more than the memory processes modeled by ACT-R's decay functions might be involved. Alternative executive-attentional measures are commonly used to study individual differences and have been shown to have high predictive power in various aspects of life (Kuncel & Hezlett, 2010). The goal of the present study is to investigate how the *rate of forgetting* relates to established measures of individual differences.

Two of the most widely used measures of individual differences are working memory capacity (WMC) and general cognitive ability or fluid intelligence (gf). There are various ways of conceptualizing and measuring both concepts. Complex span tasks are commonly used to measure WMC (Conway et al., 2005) and there are multiple standardized tests to assess general cognitive ability (e.g., the Wechsler Adult Intelligence Scale, WAIS). And while the two concepts are strongly related to each other (Ackerman, Beier, & Boyle, 2005), they are not identical (Conway, Kane, & Engle, 2003; Kane, Hambrick, & Conway, 2005).

Kane and colleagues (2007) define WMC as attentional processes that enable goal-directed behavior. Conway and colleagues (2005) recommend to administer three complex span tasks and conceptualize WMC as a composite score across those tasks. This way, task-specific components are partialled out and the derived score expresses domain general attentional processes (Kane et al., 2007). Consequently, WMC scores obtained with complex span tasks primarily reflects general executive processes (the tasks may vary but the objective is similar in other approaches, e.g., Cowan et al., 2005). Such WMC scores share variance with measures of tests of general cognitive ability because they, too, require superior executive attentional processes to obtain high scores (Engle, Tuholski, Laughlin, & Conway, 1999).

The goal of the current study is to shed light on how the *rate of forgetting* extracted from our model is related to these two measures of individual differences. A strong relationship between *rate of forgetting* and either or both WMC and general cognitive ability would suggest that the model's parameter encapsulates an executive process that is akin to what allows individuals to perform well on tests of WMC or general cognitive ability. If there were no such relationship, however, we would conclude that the *rate of forgetting* reflects a measure of individual difference that tells us something about a learner's memory retention capacities that goes beyond how well they can use their executive-attentional resources efficiently.

To this end, we report the findings from an experiment in which participants studied a set of facts so their *rate of forgetting* could be estimated. Additionally, they completed three blocks of three complex span tasks (analogous to Foster et al., 2015) as well as a test of general cognitive ability.

Methods

Procedure

All participants were invited for two sessions that were spaced three days apart.

Session 1. In the first session participants spent 20 minutes learning 35 Swahili-Dutch word-pairs. Participants were randomly assigned to study with one of two methods: either they used digital flashcards or an adaptive learning method (Sense, Behrens, Meijer, & van Rijn, 2015; Sense et al., 2016). As we will focus here on the results of the adaptive learning method, we will refrain from further discussion of the digital flashcard method.

Next, participants completed the three complex span tasks used by Foster and colleagues (2015). In these tasks, participants are shown items that need to be recalled in the correct order at the end of trial. Each to-be-remembered item is followed by a distractor, which requires the participant to engage executive attentional processes. This is to reduce the ability to rehearse to-be-remembered items during the distractor task. In the Operation Span task, for example, to-be-remembered items are letters and distractors are simple equations (e.g., $(2 \times 2) - 1 = 3$), which the participant has to made true/false judgments about. All participants completed three blocks of each task and the order of the tasks was identical across participants (as in Foster et al., 2015): first Operation Span, followed by Rotation Span, and then Symmetry Span.

Finally, a test of the word-pairs that were studied at the beginning of the session was administered. All 35 Swahili cues were shown in a list and the participant had to provide the correct Dutch translation. The test was self-paced and participants were able to provide answers in any order they preferred. No feedback was provided.

The duration of complex span tasks varies between participants. To ensure that the retention interval between the word-learning task and the test was the same across participants, a simple lexical decision task was administered as a filler task before the test. Five-letter strings were presented on screen and participants had to press one of two buttons to indicate whether the string was a Dutch word or not. By using high frequency words, the task was made relatively easy to avoid fatigue. All but two participants maintained accuracy levels above 80% and visual inspection of the response time distributions suggests the task was performed consistently. The task was chosen because it was easy to check on a trial-by-trial basis whether participants engaged in the task, and could easily be programmed to ensure that all participants started with the subsequent test at the same, relative, point in time. The data from the filler task will not be discussed further here.

Session 2. Three days later, participants came back for the second session. The second session started with a second

test of the Swahili-Dutch word-pairs learned at the beginning of the first session. The test was identical to the one completed at the end of Session 1.

Subsequently, we assessed the participant's general reasoning abilities by administering the Q1000 Cognitive Capacity test on-line. Upon completion of the test, the website provided participants with feedback indicating how their performance (overall and on the sub-scales) compared to that of a norm group.

Materials

Swahili-Dutch Word-Pairs. The 35 items were randomly sampled from the list of 100 Swahili-English word-pairs provided by Nelson and Dunlosky (1994). The English responses were translated to Dutch and all participants studied the same subset of 35 word-pairs. The order in which words were introduced was randomized.

Complex Span Tasks. The code for the three complex span task was obtained from the Engle lab's website and used with their permission. It is the same code used by Foster and colleagues (2015) but all instructions were translated to Dutch. Scores reported in Table 1 are partial-credit unit scores (Conway et al., 2005).

General Cognitive Ability. As a measure of general cognitive ability, we used *Q1000 Capaciteiten Hoog ("High Capacity"; university-educated individuals)* developed by Meurs HRM. The test has been developed for a selection context to determine whether a candidate has the necessary intellectual ability to perform well in cognitively demanding jobs, but has psychometric properties akin to other standardized tests of intelligence. There are multiple sub-scales that are ordered hierarchically with the declared goal of measuring general intelligence. In contrast to more traditional tests of general cognitive ability, this test can be administered, using online tools, in a classroom setting. The Committee on Test Affairs Netherlands (COTAN) has evaluated the test and concluded that it is a valid and reliable measure of general cognitive ability (Van Bebber, Lem, & Van Zoelen, 2010). All scores reported here are z-scores relative to the highest available norm group ("WO"; people that completed university education).

Participants

A total of 42 participants were recruited from the Dutch first-year participant pool at the University of Groningen. Of those, 14 were female (33%) and the median age was 19 (SD_{age} = 1.34; $range_{age}$ = [17, 24]). No one indicated any familiarity with Swahili. All participants gave informed consent and the Ethics Committee Psychology approved the study (ID: 15006-N).

Due to technical issues, data in the Rotation Span task was lost for one participant and in the Symmetry Span task for another. The composite scores (i.e., WMC) for these two individuals are based on the z-score average of the two remaining tasks. One participant did not complete the second vocabulary test and Q1000 scores were not available for 4 participants because extreme weather prevented the

admission of the second experimental session for one group of participants.

Results

To express a single measure of working memory capacity (WMC), the scores on the three complex span tasks are summarized into a single composite score[1]. This is done by calculating a participant's z-score for each task and then computing a z-score average (following Foster et al., 2015). Table 1 provides descriptive statistics for the scores on the individual tasks as well as their partial correlations among each other and with the resulting composite score. As expected, the complex span tasks correlate with each other and are highly correlated with the composite score. All correlation coefficients differ significantly from 0 with $p <$ .001[2]. For brevity's sake, the composite score will be referred to as a participant's WMC.

Table 1. Descriptive statistics for the complex span tasks and their composite score (WMC) as well as the correlations between all measures.

	Mean	SD	Range	1.	2.	3.
1. OSpan	59.5	9.2	[39,75]			
2. RotSpan	29.0	7.3	[10, 40]	.56		
3. Symspan	31.4	6.9	[14, 42]	.64	.54	
4. WMC	0.0	0.8	[-1.6, 1.3]	.86	.83	.85

There is considerable variation, both across items and participants, in the estimated parameters that are used to compute the *rate of forgetting*. This indicates that some words are more difficult to learn than others and that some participants leaned the material more easily than others. As described in the Introduction, the *rate of forgetting* is obtained by computing the average across all item-specific parameters estimated by the model. Across the 42 participants, the mean *rate of forgetting* is .288 with a standard deviation of 0.049 (range = [0.186; 0.404]), with higher values indicating faster forgetting. The distribution is also apparent in Figure 1 and indicates that individual differences can be captured by the *rate of forgetting*.

Figure 1 provides an overview of all relevant measures and how they relate to each other. The plot depicts the distribution of each variable on the diagonal. On the off-diagonal, scatterplots with fitted linear regression lines are shown. The corresponding partial correlations are shown on the other off-diagonal, along with *p*-values expressing the probability of observing the data assuming the coefficient is 0. Also shown are the sample size and the Bayesian

---

[1] A supplement with the raw data and scripts to compute all numbers and generate the plot in this manuscript is available at: https://github.com/fsense/cogsci-2016-paper

[2] Bayesian equivalents (Wetzels & Wagenmakers, 2012) were computed and ranged from 90 to 2327 in favor of the alternative model (that $r \neq 0$) for correlation coefficients among the three tasks and were all over one billion for the three coefficients listed in row four of Table 1.

equivalent of the null-hypothesis significance test (Wetzels & Wagenmakers, 2012). The subscript "H0" indicates that the Bayes factors quantify the evidence the data provides for the null hypothesis (assuming no correlation, i.e., that $r = 0$) relative to the alternative hypothesis and vice versa for subscript "H1".

Participants took one test of the learned word-pairs in both sessions but only the test results from the second test are included in Figure 1. The scores from the two tests are highly correlated ($r = .88$) and interchangeable[3] when it comes to the conclusions that can be drawn from Figure 1.
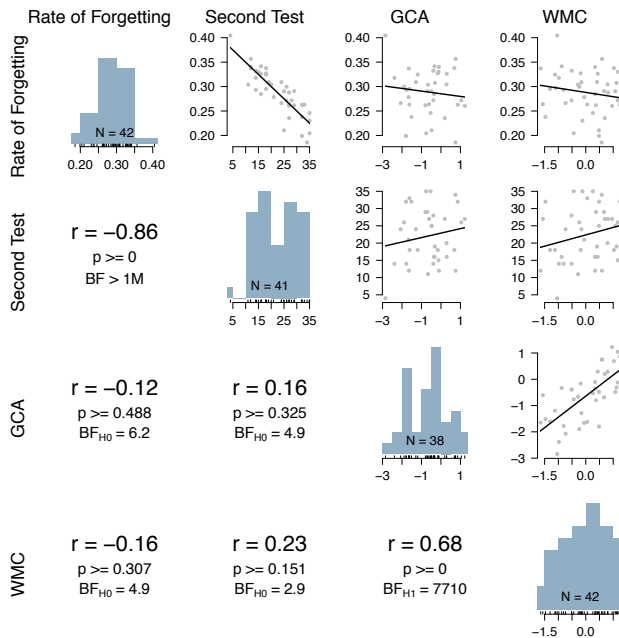

Figure 1. Depicted are the measures of interest: the estimated rate of forgetting, the score on the second test, someone general cognitive ability (GCA), and their working memory capacity (WMC).

Since a higher *rate of forgetting* indicates faster forgetting, we would expect all correlations in the left-most column of Figure 1 to be negative. This expectation is met by the performance on the second test; forgetting items faster is strongly related to performing well on the test. While the signs of the correlation coefficients corresponding to working memory capacity (WMC) and general cognitive ability (GCA) are also negative, both the *p*-values and the Bayes factors suggest that we can assume the correlations are zero. The other correlation coefficients also have the expected, positive sign but the data only provide evidence for a non-zero correlation between WMC and GCA.

Thus, the only measure that is related to test performance is the *rate of forgetting* estimated during learning, while the individual differences captured by WMC/GCA and the *rate of forgetting* do not seem to share any variance.

---

[3] See additional information provided in the Supplement for details.

Discussion

The goal of the present study is to investigate how the *rate of forgetting* estimated during the learning with our adaptive fact-learning model relates to established measures of individual differences.

The very high correlation between *rate of forgetting* and test scores (see Figure 1) suggests that the parameter estimated during learning captures the learner's ability to store and retrieve information from memory. Vocabulary tests are a common way to assess word-pair learning in many school curricula, which gives face validity to the test administered here. Being able to predict test performance with 74% accuracy (the square of $r = -.86$; see Figure 1) is not only impressive but also denotes that the estimated parameter captures meaningful individual differences.

We could neither establish a correlation between the *rate of forgetting* and someone's WMC nor their general cognitive ability. The lack of a correlation suggests that the *rate of forgetting* shares very little to no variance with the performance on either complex span tasks or a test of general cognitive ability. Since both these tasks make strong executive attention demands (Kane et al., 2007), we conclude that the *rate of forgetting* encapsulates a characteristic of the learner that is not analogous to their general cognitive ability.

Earlier work provides additional evidence for the view that the *rate of forgetting* might be independent from other cognitive span measures. For example, Rosen and Engle (1998) had people learn three lists of paired-associates. For half the participants, the second list re-used cues from the first list but participants had to associate them with different responses. This created interference and they were interested in whether low and high span participants were affected by the experimental manipulation differently. They found that low and high span participants learned the first list at the same rate, when just the *rate of forgetting* plays a role*,* but that low span participants' performance suffered a lot more from the interference than high span participants'. Their experiments suggest that high span participants perform better in the interference condition because they ward off the intrusions more successfully (also see Brewin & Beaton, 2002; however, see Oberauer, Lange, & Engle, 2004 for contradicting evidence). No such interference existed in our experiment and the finding that someone's WMC is not linked to their *rate of forgetting* is consistent with the finding that low and high span participants learn paired-associates at the same rate (Rosen & Engle, 1998). Similarly, Kane and colleagues conclude that WMC measured with complex span tasks "does not predict variability in *all* aspects of remembering" (2007, p.32) even though there is research that suggests an influence of WMC on memory-related processes (e.g., Mall & Morey, 2013).

Although the Q1000 is not a commonly used questionnaire to assess general cognitive ability, this study provides further, indirect, validation. Working memory capacity (WMC) and general cognitive ability are known to be related concepts, and in this study we find a correlation

in the expected range (Ackerman et al., 2005). Nevertheless, it would be good to compare the current findings with more commonly used tests of general cognitive ability (e.g., WAIS or Raven's advanced progressive matrices).

Based on the data presented here, we can extrapolate that the individual differences captured by someone's *rate of forgetting* are not confounded by their general cognitive abilities. We believe, however, that this relationship might emerge if a more heterogeneous sample is used. Our participants were first-year psychology students at a Dutch university, which necessarily restricts the range of general cognitive ability in an absolute sense. Figure 1 shows that there is a reasonable amount of variation in the data when it comes to general cognitive ability but this is due to the fact that the test we used is highly sensitive in the higher range of the construct it measures. Therefore, it is not unreasonable to assume that a relationship between someone's ability to learn factual knowledge and their general cognitive ability might exist in the general population. These results suggest, however, that in more restricted ranges as observed in the population commonly used in psychological research, WMC and general cognitive ability do not influence fact learning.

Conclusion

The data presented here confirm that the *rate of forgetting* estimated by the adaptive fact-learning model developed in our lab is an excellent predictor of outcomes on tests of the same material. While the common measures of individual differences used in this study - working memory capacity and fluid intelligence - are related to each other in the way we would expect, they do not seem to be related to the estimated *rate of forgetting*. This is interesting because it suggests that the rate with which someone acquires and retains factual knowledge is not linked to their executive attentional capabilities.

Acknowledgments

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: the same or different constructs? *Psychological Bulletin*, *131*(1), 30–60. doi:10.1037/0033-2909.131.1.30

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York: Oxford University Press.

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An Integrated Theory of List Memory. *Journal of Memory and Language*, *38*(4), 341–380.

Brewin, C. R., & Beaton, A. (2002). Thought suppression, intelligence, and working memory capacity. *Behaviour Research and Therapy*, *40*, 923–930. doi:10.1016/S0005-7967(01)00127-9

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: a temporal ridgeline of optimal retention. *Psychological Science*, *19*(11), 1095–102. doi:10.1111/j.1467-9280.2008.02209.x

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. doi:10.3758/BF03196772

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–552. doi:10.1016/j.tics.2003.10.005

Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*(1), 42–100. doi:10.1016/j.cogpsych.2004.12.001

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist*, *43*(8), 627–634.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, *84*(5), 795–805. doi:10.1037//0021-9010.84.5.795

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology. General*, *128*(3), 309–331. doi:10.1037/0096-3445.128.3.309

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, *43*(2), 226–236. doi:10.3758/s13421-014-0461-7

Kane, M. J., Conway, A. R. A., Hambrick, David Z, & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, & A. T. Miyake (Eds.), *Variation in working memory* (pp. 21–46). New York, NY: Oxford University Press. doi:10.1093/acprof

Kane, M. J., Hambrick, D. Z., & Conway, A. R. a. (2005). Working memory capacity and fluid intelligence are strongly related constructs: comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 66–71; author reply 72–75. doi:10.1037/0033-2909.131.1.66

Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science*, *315*, 1080–1081. doi:10.1126/science.1136618

Kuncel, N. R., & Hezlett, S. A. (2010). Fact and Fiction in Cognitive Ability Testing for Admissions and Hiring Decisions. *Current Directions in Psychological Science*, *19*(6), 339–345. doi:10.1177/0963721410389459

Mall, J. T., & Morey, C. C. (2013). High working memory capacity predicts less retrieval induced forgetting. *PloS One*, *8*(1), e52806. doi:10.1371/journal.pone.0052806

Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. *Memory*, *2*(3), 325–335.

Nijboer, M. (2011). *Optimal fact learning: Applying presentation scheduling to realistic conditions*. University of Groningen.

Oberauer, K., Lange, E., & Engle, R. W. (2004). Working memory capacity and resistance to interference. *Journal of Memory and Language*, *51*(1), 80–96. doi:10.1016/j.jml.2004.03.003

Pavlik, P. I., & Anderson, J. R. (2003). An ACT-R model of the spacing effect. In *Proceedings of the 5th International Conference on Cognitive Modeling* (pp. 177–182). Bamberg.

Pavlik, P. I., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science*, *29*(4), 559–86. doi:10.1207/s15516709cog0000_14

Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied*, *14*(2), 101–17. doi:10.1037/1076-898X.14.2.101

Rosen, V. M., & Engle, R. W. (1998). Working memory capacity and suppression. *Journal of Memory and Language*, *39*, 418–436. doi:10.1006/jmla.1998.2590

Sense, F., Behrens, F., Meijer, R. R., & van Rijn, H. (2015). Stability of Individual Parameters in a Model of Optimal Fact Learning. In *Proceedings of the 13th International Conference on Cognitive Modeling* (pp. 136–141).

Sense, F., Behrens, F., Meijer, R. R., & Van Rijn, H. (2016). An Individual's Rate of Forgetting is Stable over Time, but Differs Across Materials. *Topics in Cognitive Science*, 8(1), 305-21. doi:10.1111/tops.12183

Van Bebber, J., Lem, J., & Van Zoelen, L. (2010). *Q1000 Capaciteiten Hoog*. Woerden: Meurs HRM.

Van Rijn, H., van Maanen, L., & van Woudenberg, M. (2009). Passing the test: Improving learning gains by balancing spacing and testing effects. In *Proceedings of the 9th International Conference on Cognitive Modeling* (pp. 110–115).

Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, *19*, 1057–1064. doi:10.3758/s13423-012-0295-x