Psychological Bulletin
2005, Vol. 131, No. 1, 61–65

0033-2909/05/$12.00 DOI: 10.1037/0033-2909.131.1.61

Working Memory and Intelligence—Their Correlation and Their Relation: Comment on Ackerman, Beier, and Boyle (2005)

Klaus Oberauer
University of Potsdam

Ralf Schulze
Westfälische Wilhelms-Universität Münster

Oliver Wilhelm
Humboldt-University Berlin

Heinz-Martin Süß
University of Magdeburg

On the basis of a meta-analysis of pairwise correlations between working memory tasks and cognitive ability measures, P. L. Ackerman, M. E. Beier, and M. O. Boyle (2005) claimed that working memory capacity (WMC) shares less than 25% of its variance with general intelligence (*g*) and with reasoning ability. In this comment, the authors argue that this is an underestimation because of several methodological shortcomings and biases. A reanalysis of the data reported in Ackerman et al. using the correct statistical procedures demonstrates that *g* and WMC are very highly correlated. On a conceptual level, the authors point out that WMC should be regarded as an explanatory construct for intellectual abilities. Theories of working memory do not claim that WMC is isomorphic with intelligence factors but that it is a very strong predictor of reasoning ability and also predicts general fluid intelligence and *g*.

Ackerman, Beier, and Boyle (2005) are to be applauded for their heroic effort in bringing together the large set of findings on correlations between measures of working memory capacity (WMC) and cognitive ability tests. We believe, however, that their analysis and interpretation of these findings is partly flawed and partly biased. More important, we feel that their article reflects a misunderstanding of the theoretical meaning of correlations between current measures of WMC and intelligence tests or other ability tests. In our commentary, we first point out methodological problems with the meta-analysis of Ackerman et al. In the second part of this comment we discuss the theoretical role of that relation in the context of attempts to understand the nature of intelligence.

Methodological Issues: What Is the Correlation Between WMC and Intelligence?

The goal of Ackerman et al.'s (2005) meta-analysis was to arrive at estimates of the correlations between the WMC construct and various cognitive ability constructs. It is clear from their writing that they were intent on downplaying these correlations. We show that their procedure involves factors that bias the estimated correlations downward and results in confidence intervals that are too narrow.

Task Selection

One such factor is the selection of tasks to represent WMC. The inclusion criteria Ackerman et al. (2005) used are unclear. The text suggests that working memory (WM) tasks are dual-task paradigms that combine a short-term storage task with a processing task. The list in the Appendix of their article, however, includes tasks that do not match this description (e.g., random generation, Star Counting Test). It appears as if Ackerman et al. have included every task that the original authors have labeled a WM task. There is probably no other impartial selection criterion, so Ackerman et al. arguably had no choice but to select tasks by their labels. The consequence, however, is that the meta-analysis ignores the progress in honing the construct WMC over the last 2 decades. For example, work by our group (Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000) included random generation and the Star Counting Test in one study to test their construct validity. It turned out that these two tasks were questionable indicators of WMC—random generation shared little variance with other WM tasks, and the Star Counting Test loaded on a separate factor that reflected a mixture of executive functions (i.e., supervision of basic processes) and processing speed. Later research confirmed that executive functions form one or several factors that are not strongly related to measures of WMC (Oberauer, Süß, Wilhelm, & Wittmann, 2003). Moreover, we learned that factors reflecting WMC were excellent predictors of reasoning ability, whereas the factor reflecting speed and executive functions was a good predictor of psychometric speed but not reasoning (Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). Hence, we have a more refined picture now of WMC as a construct, so that in retrospect we have reasons to exclude tasks that do not adequately measure the construct, in particular tasks designed to measure executive functions.

Klaus Oberauer, Department of Psychology, University of Potsdam, Potsdam, Germany; Ralf Schulze, Department of Psychology, Westfälische Wilhelms-Universität Münster, Münster, Germany; Oliver Wilhelm, Department of Psychology, Humboldt-University Berlin, Berlin, Germany; Heinz-Martin Süß, Department of Psychology, University of Magdeburg, Magdeburg, Germany.

This work was in part supported by Deutsche Forschungsgemeinschaft Grant OB 121/3-3.

Correspondence concerning this article should be addressed to Klaus Oberauer, University of Potsdam, Allgemeine Psychologie I, PO Box 60 15 53, 14415 Potsdam, Germany. E-mail: ko@rz.uni-potsdam.de

By including such tasks in the meta-analysis, Ackerman et al. biased their estimate of the correlation of WMC and reasoning downward and their estimate of the correlation of WMC and speed upward.

Meta-Analysis

A second set of problems arises with the model and procedures used to conduct the meta-analysis. Ackerman et al. (2005) assumed a fixed-effects model. The assumption made by choosing this model is that all synthesized studies estimate the same constant correlation in the universe of studies. This assumption is highly questionable in general (National Research Council, 1992), and in particular for the given situation. As we discuss in the next section, the various WM tasks differ in the composition of sources of variance they reflect. Hence, their correlations with ability tests are heterogeneous in the universe of studies. In addition, the samples of participants come from different populations, and other study characteristics probably add to the heterogeneity of effect sizes. A random-effects model would have been much more adequate for this situation. The consequences of using a fixed-effects model when a random-effects model would be correct are manifold (for an overview, see Schulze, 2004). Although it is unlikely that the mean effect size estimates reported by Ackerman et al. are biased because of the use of fixed-effects model procedures, the confidence intervals are too narrow when heterogeneity prevails in the universe of studies (cf. Hedges & Vevea, 1998; Schulze, 2004). In addition, the confidence interval widths are also underestimated because the procedure used to estimate them does not take the correction for attenuation into account. Confidence intervals for corrected correlations are larger than those for uncorrected correlations (Hunter & Schmidt, 1990). As a consequence, some conclusions drawn by Ackerman et al. based on nonoverlapping confidence intervals might have been different when using appropriate methods. For example, the average correlation between elementary cognitive task and WM tests might not significantly exceed the average correlation between WM tests and general intelligence (g) measures.

Unwanted Variance and Task Specificity

Another problem leading to an underestimation of the correlation between WMC and cognitive abilities is that no single task used to measure WMC is a pure indicator. WM tasks arguably reflect a mix of at least four sources of systematic variance: (a) variance of the construct WMC, (b) variance specific to the task paradigm (e.g., dual tasks of storage and processing, running memory task, short-term recognition task, relational integration tasks), (c) content-related variance, and (d) method variance (e.g., computer-based assessment, time limits of administration). The latter three are unwanted sources of variance that usually diminish the measured correlation with criterion variables (i.e., if they do not affect the criterion variable as well). Because they are systematic sources of variance, their effect is not corrected by the correction for attenuation that Ackerman et al. (2005) applied.

One way to limit the impact of unwanted variance is to assess WMC with a battery of tasks selected to balance different task paradigms, contents, and methods, ideally approaching representative coverage of the universe of possible WMC tests. Under these conditions, unwanted variance can be averaged out through aggregation. There are a few studies in the literature that come relatively close to that goal (Kane et al., 2004; Kyllonen, 1994; Süß, Oberauer, Wilhelm, & Wittmann, 2000; Süß et al., 2002). It is interesting to note that in three of these four studies WM factors accounted for considerably more variance in reasoning ability or general fluid intelligence (g_f) than estimated by Ackerman et al. (2005; the exception being Kane et al., 2004, who used only storage-and-processing tasks to measure WMC). In a reanalysis of the two studies by Süß et al. (2000, 2002), Oberauer, Süß, Wilhelm, and Sander (in press) demonstrated how the correlation obtained between measures of WMC on the one side, reasoning and general intelligence on the other side increased as the level of aggregation on the WMC side was increased. Individual tasks such as reading span or computation span had correlations with the intelligence scales in the range reported by Ackerman et al. (rs = .24 to .64). At a first level of aggregation Oberauer et al. (in press) formed composites of tasks measuring the same WM function (simultaneous storage and processing or relational integration) with the same content (verbal–numerical or spatial). These composites had correlations with intelligence ranging from .50 to .69. At a second level of aggregation, the authors built a single WMC composite representing both functions and both content domains. This composite correlated with intelligence scales .69 to .77.

Ackerman et al. (2005) are aware of the merits of aggregation, but they largely disregarded it in their assessment of the correlation between WMC and intellectual abilities, which they based largely on estimates of correlations between individual WM tasks and ability test scores. Whereas the hypothesis they intended to test was formulated on the level of constructs, their data analysis focused on the level of individual indicators and thereby fell short of testing the hypothesis.

Structural Equation Models

A more sophisticated procedure than using aggregation of indicators is the application of structural equation modeling (SEM). It can be used to estimate the strength of relationship between latent variables that assumedly represent the constructs of interest. Latent variables (e.g., WMC and g_f) figure as causes to explain the covariances between their indicators (WM tasks and ability test scores). Having covariances between all indicators available and attending to certain restrictions when applying SEM enables the estimation of the latent variables' relationship.

Ackerman et al. (2005) attempted to apply this approach to a subset of uncorrected correlation estimates from their meta-analysis. They seemed to be aware of some serious statistical problems with such an analysis, but they conducted their analyses in a way that disregards one of these problems, one which is highly relevant when using correlation matrices as data input for SEM. The standard procedures (and software) for SEM are designed for application to covariance matrices, not correlation matrices. Unless so-called *scale-invariant* models and *scale-free* parameters are used, omnibus test statistics, fit indices, as well as parameter estimates and their standard errors, may be severely biased (Cudeck, 1989) when correlation matrices are modeled. The critical test of the hypothesis that WMC and g are isomorphic constructs is the model in Ackerman et al.'s Figure 2. They chose to fix the loadings of the indicators in that model on the basis of their loadings in a different model. Apart from being theoretically unjustified, this decision has the consequence that the model is not scale invariant.

We therefore reanalyzed the data in Ackerman et al.'s (2005) Table 4, fitting their model with unconstrained factor loadings and using software implementing a constrained estimation algorithm to properly analyze correlation matrices.[1] In addition, we changed the correlation between the WMC and the *g* factor into a unidirectional path, reflecting our conceptual stance discussed below (in this model, the standardized parameter estimate is identical to the correlation between the latent variables). The model is depicted in Figure 1. The fit indices are as follows: $\chi^2(76, N = 114) = 114.218$, $p = .003$; normed fit index = .82, comparative fit index = .93, root-mean-square error of approximation = .057, 90% confidence interval = .025, .083.

The model resulted in a substantially higher estimated correlation between WMC and *g* than that reported by Ackerman et al. (2005; $r = .85$, as opposed to $r = .50$). The variance of the disturbance term shows that the WMC factor does not account for all the variance in the *g* factor. Hence, we might conclude—setting aside the above mentioned caveats for such analyses—that WMC and *g* share the largest part of their variance (72%) but are not identical. This result converges closely with previous studies based on SEM (Kane et al., 2004; Kyllonen & Christal, 1990; Süß et al., 2002). We also reanalyzed the model relating *g* and short-term memory (Ackerman et al.'s, 2005, Figure 4) with free loadings. The path coefficient was .48, very close to Ackerman et al.'s estimate. The reanalysis therefore also resolves the conflict between Ackerman et al.'s synthesis of correlations, which confirmed that WM tasks are better predictors of intellectual abilities than short-term memory tasks, and their SEM, which did not reflect that difference.

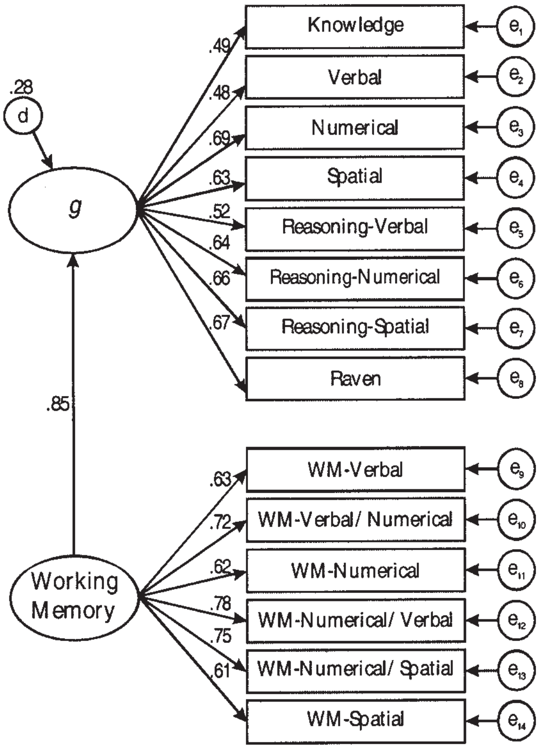

Figure 1. Reanalysis of the structural equation model in Figure 2 of Ackerman et al. (2005), based on their Table 4, with free loadings. The parameter estimates are given for the standardized solution in which constrained estimation forces endogenous latent variables to have variances of one. WM = working memory; d = the disturbance term; e = error.

Our methodological critique notwithstanding, we believe that Ackerman et al. (2005) are right in claiming that WMC is not the same as *g* or as g_f or as reasoning ability. Our argument for a distinction between these constructs does not hinge on the size of the correlation but on a qualitative difference: On the side of intelligence, there is a clear factorial distinction between verbal and numerical abilities (e.g., Süß et al., 2002); on the side of WMC, tasks with verbal contents and tasks with numerical contents invariably load on the same factor (Kyllonen & Christal, 1990; Oberauer et al., 2000). This mismatch between WMC and intelligence constructs not only reveals that they must not be identified but also provides a hint as to what makes them different. We think that verbal reasoning differs from numerical reasoning in terms of the knowledge structures on which they are based: Verbal reasoning involves syntax and semantic relations between natural concepts, whereas numerical reasoning involves knowledge of mathematical concepts. WMC, in contrast, does not rely on conceptual structures; it is a part of the architecture that provides cognitive functions independent of the knowledge to which they are applied. Tasks used to measure WMC reflect this assumption in that researchers minimize their demand on knowledge, although they are bound to never fully succeed in that regard. Still, the minimization works well enough to allow verbal and numerical WM tasks to load substantially on a common factor. This suggests that WMC tests come closer to measuring a feature of the cognitive architecture than do intelligence tests.

[1] We have used the module SEPATH included in the STATISTICA software package (Version 6.1) for the reanalyses. We noted two oddities in the data reported by Ackerman et al. (2005). First, the submatrices for the correlations between the intelligence tests differ between Tables 4 and 6. For example, the correlation between Spatial Intelligence and Knowledge was imputed in Table 4 ($r = .236$) but was seemingly based on available data in Table 6 ($r = .430$). Despite such discrepancies, we used the values in Table 4 as reported to maintain comparability. Second, we note that the harmonic mean of the sample size ($N = 456$) is much larger than the arithmetic mean for the total sample of studies ($N = 114$). It is easy to prove that the harmonic mean must be smaller than the arithmetic mean for positive values, as is the case for sample sizes. Because the sample sizes reported by Ackerman et al. for the various meta-analytic results suggest that the arithmetic mean of 114 is the more adequate value, and because the choice of which type of mean one uses for the SEM analyses does not affect the parameter estimates, the fit indices for our reanalyses reported in the text are based on an *N* of 114. The fit indices in which *N* equals 456 are as follows: $\chi^2(76) = 459.905$, $p < .001$; normed fit index = .82, comparative fit index = .85, root-mean-square error of approximation = .100, 90% confidence interval = .090, .109. Allowing correlated errors, as Ackerman et al. did, improved the fit and slightly increased the path from WMC to *g*.

Conceptual Issues: What Is the Relation Between WMC and Intelligence?

Ackerman et al. (2005) treated WMC as one more beast in the zoo of ability constructs. They were content with giving it its place in the three-stratum theory of Carroll—with an inclination toward relegating it into the rank and file, together with lower level constructs such as psychometric speed. We think that this reflects a misunderstanding of why most researchers are interested in the correlation between WMC and intelligence. The aim of that research is to validate WMC as an explanatory construct for intellectual abilities. The psychometric ability constructs have been derived largely inductively, reflecting the common variance among tests that have been constructed as diagnostic tools for aspects of mental abilities as described in everyday language. In contrast, WMC is a construct that derives deductively from theories of the cognitive architecture in which a limited-capacity WM plays a central role, although not always under the same name (Anderson & Lebiere, 1998; Atkinson & Shiffrin, 1968, to cite just the most prominent ones). These theories assign short-term memory or WM a crucial role for complex tasks such as reasoning and text comprehension. The search for correlations between measures of short-term memory (e.g., Perfetti & Goldman, 1976), and later of WM (Case, Kurland, & Goldberg, 1982; Daneman & Carpenter, 1980), and standardized ability and achievement tests has been undertaken as a test of these theories. The prediction to test was that measures of WMC should substantially correlate with the criterion tests, not that they should be perfectly correlated. Occasionally, researchers have been impressed by the size of the correlations obtained, and some have gone as far as speculating that WMC accounts for all the systematic variance in *g*, as the citations gathered by Ackerman et al. document. The scarcity of citations claiming that WMC and *g* (or g_f) are identical underscores that this is not a commonly held assumption.

From a theoretical point of view, there is no reason to assume that WMC is the same as *g*. By definition, *g* is conceptually opaque—it is the common variance of a set of tasks that happened to be constructed and used by intelligence researchers over a century. It reflects no explicit theoretical concept, and hence there is no theory-based procedure for measuring it. Rather, *g* reflects a mixture of the mostly implicit theories of intelligence various researchers have endorsed and their intuitions about ways to test it. It would be a surprise and an embarrassment if one found that measures of WMC and measures of *g* were perfectly correlated. It would imply that measures of WMC do not come closer to measuring a theoretically well-defined parameter of the cognitive system than *g* does.

By treating WMC as another primary factor in the ability hierarchy, Ackerman et al. (2005) ignore its theoretical background. WMC is a construct that bridges the gap between research on individual differences in abilities and cognitive science, including experimental cognitive psychology and formal modeling of cognitive processes. The tasks used to measure WMC have been constructed to operationalize processes postulated in theories of WM, and although these theories are admittedly still in their infancy, they provide some guidance as to what features a good WM task should have. More important, the link between theory and measurement implies that the theory can be tested through construct validation of the tests (Oberauer, in press). Investigating the correlates of WM tests is one aspect of construct validation; these efforts are complemented by experimental investigations, investigations of neural processes, and formal modeling. A wealth of experimental findings informs about the processes going on in the most popular WM tasks (e.g., LaPointe & Engle, 1990; Saito & Miyake, 2004), and neuroimaging studies unravel the networks involved in these processes (Curtis & D'Esposito, 2003). Moreover, there are encouraging attempts to develop formal models of the capacity limit of WM (Daily, Lovett, & Reder, 2001; Oberauer & Kliegl, 2001). The field is still far from consensus about what WM is and how it functions, but there are various competing theories that make testable empirical claims (for an overview, see Miyake & Shah, 1999). Nothing comparable can be said about *g*.

We want to understand intelligence, not only map its network of correlations with other constructs. This means to reveal the functional—and ultimately, the neural—mechanisms underlying intelligent information processing. Among the theoretical constructs within current theories of information processing, WMC is the one parameter that correlates best with measures of reasoning ability, and even with g_f and *g*. Therefore, investigating WMC, and its relationship with intelligence, is psychology's best hope to date to understand intelligence. Stopping short at searching for the place of WMC among the factor hierarchy of ability constructs is like being satisfied with a Linnéan taxonomy of creatures and refusing to proceed toward explaining the origin of species.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131,* 30–60.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Vol. 2. Advances in research and theory* (pp. 90–195). New York: Academic Press.

Case, R., Kurland, M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33,* 386–404.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin, 105,* 317–327.

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences, 7,* 415–423.

Daily, L. Z., Lovett, M. C., & Reder, L. M. (2001). Modeling individual differences in working memory performance: a source activation account. *Cognitive Science, 25,* 315–353.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3,* 486–504.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. E. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133,* 189–217.

Kyllonen, P. C. (1994). Aptitude testing inspired by information processing: A test of the four-sources model. *Journal of General Psychology, 120,* 375–405.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity?! *Intelligence, 14,* 389–433.

LaPointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1118–1133.

Miyake, A., & Shah, P. (1999). *Models of working memory. Mechanisms of active maintenance and executive control*. Cambridge, England: Cambridge University Press.

National Research Council. (1992). *Combining information: Statistical issues and opportunities for research*. Washington, DC: National Academy Press.

Oberauer, K. (in press). The measurement of working memory capacity. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence*. Thousand Oaks, CA: Sage.

Oberauer, K., & Kliegl, R. (2001). Beyond resources: Formal models of complexity effects and age differences in working memory. *European Journal of Cognitive Psychology, 13,* 187–215.

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—Facets of a cognitive ability construct. *Personality and Individual Differences, 29,* 1017–1045.

Oberauer, K., Süß, H.-M., Wilhelm, O., & Sander, N. (in press). Individual differences in working memory capacity and reasoning ability. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory*. New York: Oxford University Press.

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittmann, W. W. (2003). The multiple faces of working memory—Storage, processing, supervision, and coordination. *Intelligence, 31,* 167–193.

Perfetti, C. A., & Goldman, S. R. (1976). Discourse memory and reading comprehension skill. *Journal of Verbal Learning and Verbal Behavior, 15,* 33–42.

Saito, S., & Miyake, A. (2004). An evaluation of the task switching account of working memory span scores: Evidence against a temporal decay assumption. *Journal of Memory and Language, 50,* 425–443.

Schulze, R. (2004). *Meta-analysis: A comparison of approaches*. Cambridge, MA: Hogrefe & Huber.

Süß, H.-M., Oberauer, K., Wilhelm, O., & Wittmann, W. W. (2000, July). *Can working memory capacity explain reasoning ability?* Paper presented at the 27th International Congress of Psychology, Stockholm, Sweden.

Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working memory capacity explains reasoning ability—and a little bit more. *Intelligence, 30,* 261–288.

Received July 13, 2004
Revision received July 15, 2004
Accepted July 28, 2004 ■