Rapid Automatized Naming and Reading Performance: A Meta-Analysis

Susana Araújo
University of Algarve and University of Lisbon

Alexandra Reis
University of Algarve

Karl Magnus Petersson
University of Algarve; Max Planck Institute for
Psycholinguistics, Nijmegen, the Netherlands;
and Radboud University

Luís Faísca
University of Algarve

Evidence that rapid naming skill is associated with reading ability has become increasingly prevalent in recent years. However, there is considerable variation in the literature concerning the magnitude of this relationship. The objective of the present study was to provide a comprehensive analysis of the evidence on the relationship between rapid automatized naming (RAN) and reading performance. To this end, we conducted a meta-analysis of the correlational relationship between these 2 constructs to (a) determine the overall strength of the RAN–reading association and (b) identify variables that systematically moderate this relationship. A random-effects model analysis of data from 137 studies (857 effect sizes; 28,826 participants) indicated a moderate-to-strong relationship between RAN and reading performance ($r = .43$, $I^2 = 68.40$). Further analyses revealed that RAN contributes to the 4 measures of reading (word reading, text reading, non-word reading, and reading comprehension), but higher coefficients emerged in favor of real word reading and text reading. RAN stimulus type and type of reading score were the factors with the greatest moderator effect on the magnitude of the RAN–reading relationship. The consistency of orthography and the subjects' grade level were also found to impact this relationship, although the effect was contingent on reading outcome. It was less evident whether the subjects' reading proficiency played a role in the relationship. Implications for future studies are discussed.

Keywords: reading, rapid naming, moderators, meta-analysis

Supplemental materials: http://dx.doi.org/10.1037/edu0000006.supp

The cognitive correlates of reading development and reading skill are timely and important research topics because they help in grasping the nature of the cognitive processes underlying reading. Identifying these cognitive correlates may prove critical in delineating the longitudinal predictors of reading skill and may ultimately facilitate better predictions for reading difficulties. Along with phonological awareness, the ability to identify and manipulate units of spoken sound, a basic cognitive skill that has frequently been associated with reading ability is the rapid, automatic naming of visual items (for a review, see Kirby, Roth, Desrochers, & Lai,

2008). Rapid naming, also known as rapid automatized naming (RAN), refers to the time required for a child to quickly and accurately name an array of well-known visual stimuli (usually letters, digits, objects, or colors). Norton and Wolf (2011) recently argued that RAN constitutes "a microcosm or mini-circuit of the later-developing reading circuitry" (p. 430), appealing to the fact that both systems involve closely related cognitive processes. RAN tasks and reading are thought to require (a) attention to the stimuli; (b) visual processes that are responsible for initial feature detection, visual discrimination, and letter/letter-pattern identification;

(c) integration of visual information with stored orthographic and phonological representations; (d) lexical processes, including access and retrieval of phonological codes; and (e) organization of articulatory output (Wolf & Bowers, 1999).

A substantial body of evidence currently supports RAN as one of the best—and perhaps one of the most universal, longitudinal, and concurrent—predictors of reading ability and also a core deficit in reading disability (e.g., Araújo et al., 2011; Kirby, Parrila, & Pfeiffer, 2003; Kirby et al., 2008; Vaessen et al., 2010; Wolf, Bowers, & Biddle, 2000; Wolf et al., 2002). Nevertheless, there is considerable variation in the literature on the magnitude of the relationship between RAN and reading performance. Some studies have reported a strong correlation between measures of RAN and reading (e.g., Babayiğit & Stainthorp, 2010), whereas others have found small (e.g., Cunningham, 2006) or non-significant (e.g., Scarborough, 1998) correlations. Factors that varied between studies, including the methods used to assess RAN and reading performance and the characteristics of the populations studied (e.g., the participants' reading level and age range), may prove important in understanding these discrepancies. Unfortunately, the data concerning the impact of these factors are largely lacking. In a meta-analysis, Swanson, Trainin, Necoechea, and Hammill (2003) addressed the correlations between phonological awareness, RAN, and reading and found that phonological awareness and RAN were moderately correlated with word reading (.48 and .46, respectively). However, in this meta-analysis only those articles that report measures for all three constructs were included. Restricting the inclusion criteria to this precluded many suitable studies on the relation between RAN and reading, and, therefore, their findings may not generalize to the larger universe of studies addressing the RAN–reading association. More important, such restriction on the sample of available studies (only 35 studies were included) implied that the number of independent samples was not large enough to allow for testing potentially important moderator variables (e.g., format of rapid naming task).

The objective of the present study was to provide a comprehensive analysis of the evidence on the relationship between RAN and reading performance, extending the previous one in a way that we could fine-tune the analysis to several categories within relevant moderator variables. To this end, we conducted a meta-analysis of the correlational evidence on the relationship between these two variables to (a) determine the overall strength of the RAN-reading association and (b) identify variables that systematically moderate this relationship and may explain the variation between studies. The present study aimed to extend the previous meta-analysis at least in four important respects:

1. To test whether variability across types of dependent measures moderates the relation of RAN to reading, taking into account how reading skills were assessed (e.g., scoring method) as well as the characteristics of the RAN tasks (e.g., stimulus type, resource demands). The scope of Swanson et al. (2003) in reading ability was narrowly confined to real-word reading, whereas in our meta-analysis we also tested the impact of the reading domain assessed (word and text reading, non-word reading, and reading comprehension).

2. To use a more stringent criterion to define groups based on reading skill. We distinguished between poor and average readers, or a combination of both, and also considered randomly selected

samples. The distinction for this group is relevant, as it corresponds most to representative population samples.

3. To examine whether the RAN association evolves differently over time for different reading scoring methods. Swanson et al. (2003) found that correlations do not vary by age, but in this meta-analysis reading scores were pooled together, and this factor may be a confounder, as suggested elsewhere.

4. To test the moderator role of orthographic consistency and the writing system. This is a timely and important question on the grounds of recent evidence showing that script transparency might impact the cognitive dynamics of reading development.

Several findings of the present meta-analysis may also have implications for the main theories attempting to account for the relationship between RAN and reading. To date, there is still no consensus regarding the mechanisms responsible for this relationship. While some researchers have suggested that RAN primarily reflects the access and retrieval of phonological codes from long-term memory (e.g., Chiappe, Stringer, Siegel, & Stanovich, 2002; Pennington, Cardoso-Martins, Green, & Lefly, 2001; Schatschneider, Carlson, Francis, Foorman, & Fletcher, 2002; Torgesen, Wagner, & Rashotte, 1994; Wagner, Torgesen, Laughon, Simmons, & Rashotte, 1993), others have viewed RAN as an independent non-phonological reading process (Wolf & Bowers, 1999). In accordance with the latter position, Bowers and colleagues proposed an orthographic, rather than phonological, basis for the relation between RAN and reading (Bowers & Newby-Clark, 2002; Bowers & Wolf, 1993; Wolf et al., 2000). This hypothesis has since received some empirical support (e.g., Araújo et al., 2011; Georgiou, Parrila, Kirby, & Stephenson, 2008; Roman, Kirby, Parrila, Wade-Woolley, & Deacon, 2009), although others have disputed this view (Cutting & Denckla, 2001; Moll, Fussenegger, Willburger, & Landerl, 2009; Papadopoulos, Georgiou, & Kendeou, 2009).

To gain insight into the heterogeneity observed in the reported correlations between RAN and reading performance, we next review the most relevant factors that have varied between studies.

Reading Tasks

Domain of Reading

Several studies have reported that RAN is more strongly related to literacy tasks that rely on the recognition of word-specific orthographic patterns (such as the reading of irregular words or high-frequency words) compared to tasks requiring phonological decoding skills (such as non-word reading; e.g., Araújo et al., 2011; Clarke, Hulme, & Snowling, 2005; Georgiou, Parrila, Kirby, & Stephenson, 2008; Pennington et al., 2001; Savage et al., 2005). These findings suggest the hypothesis that RAN may reflect a process involved in orthographic processing, although the exact nature of this underlying process is still unknown. For example, Bowers and colleagues (Bowers & Wolf, 1993; Bowers & Newby-Clark, 2002) proposed that slow processing speed as captured by RAN tasks prevents the precise integration of visual letter sequence information in words, which is necessary to identify commonly occurring orthographic patterns, and therefore hinders the efficient acquisition of an orthographic lexicon. However, other studies have found that RAN related equally well to word and non-word reading (Moll et al., 2009; Savage, Pillay, & Melidona,

2007), which is difficult to reconcile with a pure orthographic explanation. In this study, we sought to clarify whether there is a differential relation of RAN to different dimensions of reading and to clarify how the variability in tasks used to assess reading may partially account for the variation in RAN–reading performance correlations. If the association between RAN and reading is mediated by factors related to or subserving orthographic skills and orthographic skill development (Bowers & Newby-Clark, 2002; Bowers & Wolf, 1993), then we would expect the relationship to be stronger for reading tasks that rely more heavily on orthographic knowledge.

Method of Assessing Reading

Studies have also differed in whether the outcome used to assess reading performance is accuracy-based or fluency-based. There are some indications that the RAN–reading speed association is stronger than the RAN–reading accuracy association (Georgiou, Parrila, & Kirby, 2009; Savage & Frederickson, 2005), although some exceptions have been reported in which RAN was similarly related to both accuracy and fluency (see Georgiou, Parrila, & Liao, 2008). Similarly, it is not completely clear whether the developmental relationship between RAN and reading is affected by the type of reading score. For example, a closer look at Torgesen, Wagner, Rashotte, Burgess, and Hecht's (1997) results shows that the correlations between RAN and reading accuracy drop across time, but the correlations between RAN and reading fluency do not. A similar finding was reported by Georgiou, Parrila, Kirby, and Stephenson (2008), whereas Kirby and colleagues (2003) found that the contribution of RAN to reading accuracy increases over time. One of the goals in this study was to compare the strength of the relationship between RAN and both reading accuracy and fluency in children at different school grades.

RAN Tasks

RAN Stimulus

Another source of between-studies variability is that researchers have assessed RAN performance using different types of stimuli. Some studies have reported that RAN performance on non-alphabetic stimuli, such as colors or pictured objects, before the subjects had entered school is a predictor of later reading development (de Jong & van der Leij, 1999; Kirby et al., 2003; Landerl & Wimmer, 2008). After initial literacy development, and with increasing exposure to letters and numbers, alphanumeric RAN tends to show higher correlations with reading performance than does non-alphanumeric RAN (de Jong, 2011; Lervåg & Hulme, 2009; Meyer, Wood, Hart, & Felton, 1998; van den bos, Zijlstra, & van den Broeck, 2003). For example, in a study sample of 10-year-old children, Savage and colleagues (2005) observed that the relationship between RAN and reading level was significantly higher when digit naming as opposed to object naming was used as a measure of RAN. Likewise, the results of Vaessen and Blomert (2010) revealed that RAN digits displays higher correlations with reading fluency than do RAN letters in first to sixth graders, while RAN objects display the lowest correlations. Thus, in this meta-analysis we anticipated that the relationship between RAN and

reading performance would be stronger for alphanumeric compared to non-alphanumeric stimuli.

Resource Demands of the RAN Tasks

Additional modifications to the original RAN task (Denckla & Rudel, 1976) have occurred in many experiments addressing the RAN–reading relationship. Studies have varied, for example, in the length of the tasks, diverging from the standard 50 items. In addition, in some studies the RAN tasks included a larger number of different token items than that used in the classical RAN paradigm (five token items for each subtest in the classical RAN paradigm; six in Georgiou, Parrila, Kirby, Stephenson, 2008; 20 in Cobbold, Passenger, & Terrel, 2003; and 10 or 25 for letter naming in Clarke et al., 2005). It is unknown whether these differences in task format lead to different degrees of association between RAN and reading. Norton and Wolf (2011) noted that, although small differences in RAN task length are unlikely to have an important impact as long as sufficient item familiarity is ensured, the number of tokens to be retrieved from long-term memory might be. Arguably, increasing the number of items to be accessed in RAN tasks could result in increased phonological coding. To clarify this issue, we tested whether the total length and number of tokens in the RAN task moderate the association between RAN and reading performance.

Developmental Trends

Most models of reading development assume that reading acquisition and development involves a shift in reading strategies from slow, sequential phonological decoding to fast recognition of whole-word forms (for a review, see Ehri, 2007). The dynamic nature of the reading process thus suggests that the cognitive correlates of the reading may change during literacy development.

RAN appears to be related to reading from the very start of reading acquisition (de Jong & van der Leij, 1999; Kirby et al., 2003; Landerl & Wimmer, 2008). Yet, evidence on the developmental relation between RAN and reading performance are inconsistent. On the one hand, some studies have suggested a time-limited relation, with the effect of RAN limited to the early phases of reading acquisition. For example, in the longitudinal study of Torgesen and colleagues, the correlation between RAN and reading decreased with the development of reading skill; RAN primarily influenced reading skills in Grades 1 and 2 and was not causally related to reading for Grades 3 to 5 (Torgesen et al., 1997). Similarly, Araújo and colleagues (2011) found that the association of RAN with reading outcomes is significant in normal 7-year-old readers but not in normal readers aged 9–10 years. Other studies, in turn, have rather suggested a persistent and developmentally increasing relation between RAN and reading (Vaessen et al., 2010; Vaessen & Blomert, 2010; van den bos, Zijlstra, & Spelberg, 2002; Ziegler et al., 2010). Landerl and Wimmer (2008) showed that RAN assessed at school entry (Grade 1) was a consistent predictor of reading fluency up to Grade 8.

In summary, although RAN is associated with reading development, the extent of this association and how it may change over time are not yet clear. In this study we investigated the relationship of RAN to reading accuracy and speed at different stages during reading development.

Influence of Reading Status

RAN predicts the reading level not only in normally developing readers but also in poorly developing readers, such as those with dyslexia. The relative contribution of RAN performance to predicting reading scores might even be stronger for less-proficient readers (e.g., Araújo, Pacheco, Faísca, Petersson, & Reis, 2010; Johnston & Kirby, 2006; McBride-Chang & Manis, 1996; but see, for example, Katzir et al., 2006). Moreover, slow RAN is known to distinguish dyslexics from normally developing readers, "garden-variety" poor readers, and readers with other learning disabilities (for an overview, see Denckla & Cutting, 1999; Wolf et al., 2000). Meyer and colleagues have suggested that the RAN–reading relationship is strong and persistent (at least up to eighth grade) only in poor readers (Meyer et al., 1998). McBride-Chang and Manis (1996) also found that RAN correlated with reading skill for poor readers but not for good readers. Similarly, in a more recent study of Araújo and colleagues (2010), the reading performance in dyslexic children was found to be predicted by RAN, while in age-matched normal readers it was non-significant (after controlling for IQ, vocabulary, and phonological awareness). However, these results contradict previous findings indicating that RAN was more strongly associated with word recognition and word decoding in non-impaired readers compared to dyslexic readers (Katzir et al., 2006). Note that in this study (unlike in Araújo et al., 2010), groups were matched based on word-reading proficiency, and consequently dyslexics were older than the normal readers. Tentatively, naming speed within average readers may be a transitory correlate of reading, while slow naming speed may persist as a characteristic of disabled readers. One of our goals in the present study was to test whether the strength of the relation between RAN and reading depends on the readers' proficiency and to examine whether this effect interacts with grade.

Influence of Orthographic Consistency

Alphabetic orthographies differ with respect to how consistently letters map onto their corresponding speech sounds. In writing systems referred to as *consistent* or *transparent*, such as Italian or Finish, the orthographic and phonemic codes are isomorphic (i.e., the correspondences between graphemes and phonemes are regular and unambiguous). By contrast, in writing systems referred to as *inconsistent* or *opaque*, such as English, the same grapheme is often pronounced differently in different words because the relation between graphemes and phonemes is not one-to-one (Ziegler & Goswami, 2005; Ziegler, Perry, Jacobs, & Braun, 2001). The orthographic consistency of a writing system is a key factor for determining the rate at which reading is acquired (Seymour, Aro, & Erskine, 2003) and also influences fundamental aspects of reading, such as the grain size of basic reading units (Frost, 2005; Ziegler et al., 2001).

It follows that the orthographic structure of the language itself should be reflected in the correlates of early reading acquisition. To date, the cross-linguistic comparisons addressing this topic have produced mixed results. Some studies have reported that phonological awareness is more important in learning to read in English, while rapid naming is more important for transparent orthographies, such as Dutch, Greek, or German (Georgiou, Parrila, & Papadopoulos, 2008; Landerl & Wimmer, 2008; Mann & Wimmer, 2002). Others have found that rapid naming is consis-

tently less important than phonological awareness, regardless of transparency (Caravolas, Volin, & Hulme, 2005; Patel, Snowling, & de Jong, 2004).

In interpreting findings of cross-language differences in the RAN–reading association, it is important to bear in mind one important issue. Reading skill is usually measured in terms of accuracy in opaque orthographies such as English, while it is more commonly measured by speed in transparent orthographies (owing to the ceiling effects in the reading accuracy scores). Consequently, the apparent stronger relation between RAN and reading as sometimes reported for transparent orthographies may be an artifact of shared method variance. Ziegler and colleagues (2010) recently demonstrated that when equivalent reading measures were used in five languages on a continuum of orthographic consistency (Finnish, Hungarian, Dutch, Portuguese, and French), the influence of RAN was not significantly modulated by script transparency. However, in Ziegler et al. (2010) RAN was measured using object naming only. Different results may have been obtained if the measure of RAN had involved speeded naming of letters and/or digits. In the meta-analysis in this article, we sought to extend previous studies to shed new light on whether the association between alphanumeric and non-alphanumeric RAN and reading measures is systematically modulated by script transparency.

None of the studies mentioned so far has investigated how orthographic consistency modulates the developmental course of the cognitive skills underlying reading. Vaessen and colleagues (2010) addressed this issue in a cross-sectional study with children in Grades 1 through 4 with Hungarian, Dutch, and Portuguese as their native languages. The authors found that orthographic consistency did not influence the contribution of RAN to reading fluency in any of the grades tested, suggesting that the strength of the relationship between RAN and reading is equally strong in opaque and transparent orthographies. In the current study, we tested whether the degree of regularity of grapheme–phoneme correspondences influences the time frame when the RAN–reading association is significant. At least three alternative claims have been made: (1) RAN is an important correlate of reading in all alphabetic writing systems but has varying roles based on the orthography being studied; (2) RAN is equally important in opaque and transparent orthographies; or (3) RAN is equally important in all orthographies, although not necessarily over the same time frame.

Finally, because one limitation of previous cross-language studies is their focus on alphabetic scripts, in this meta-analysis we also included systematic cross-script exploration.

Method

Literature Search Procedure and Inclusion Criteria

The studies included in the meta-analysis were identified and selected by searching the PsycINFO, PubMed, and Web of Knowledge databases. The search strategy used the following combination of terms: ("RAN OR rapid naming OR rapid automatized naming OR rapid serial naming OR naming speed") AND ("reading OR predictors of reading"). Our search covered all of the published articles that were available in the databases up until January 2014. This procedure yielded 877 citations from articles. We also checked the reference lists of

prior meta-analyses and narrative reviews on RAN and reading competence, enabling us to identify 31 additional references. We then evaluated each article to determine which studies were eligible for inclusion in the meta-analysis (46 of the articles were not written in English and therefore were not included in the evaluation). The following criteria were required for inclusion: (1) original empirical data were reported; (2) the study was based on direct tests of RAN and reading competence (not teacher/parent rating scales or surveys); (3) the sample size and zero-order correlations among measures of RAN and reading ability were reported; (4) concurrent RAN and reading assessments were conducted; and (5) the participants in the study were children or adolescents (mean sample age below 18). Thus, dissertations, unpublished studies, or studies not reporting concurrent correlations between RAN and reading (e.g., longitudinal studies reporting coefficients between tasks in Grades 1 and 2 but not correlations within grades), not using the classical RAN paradigm and direct tests of reading, or presenting data from adults only, were excluded from the present meta-analysis. Data from clinical samples other than dyslexics (e.g., Williams' syndrome) were also excluded. Finally, because including correlations based on aggregated multi-aged samples may bias the effect sizes estimate, whenever multi-age studies covered grade ranges greater than 3 years, only those for which scores on standardized tests of reading and RAN were age- or grade-corrected were considered.

To prevent violation of the independence of the observations (by including data from the same sample more than once), studies with identical authors were examined for duplicate samples. Whenever sample overlap occurred, we included the article that reported a larger or more complete data set and excluded the overlapping studies. For longitudinal studies, the first time point was coded, and for experimental studies, only pre-test data were coded. If a study reported the complete correlation matrix for composite scores as well as those of individual tests or subtests, preference was given to coding the individual scores because these variables could be important to explain variation between studies. Similarly, preference in coding was given to independent samples. For example, a study may have reported separate matrices for impaired readers and average readers, as well as a correlation matrix for both samples combined. In such cases, we discarded the matrix for the combined sample. In cases where combined matrices were presented for the moderator variables of interest, without reporting separate effect sizes, we contacted the author of the article and asked for the individual data, allowing us to retain 161 separate effect sizes associated with moderators.

Of the initial 908 articles, only 151 articles representing 165 studies met all of our meta-analysis inclusion criteria. All of the selected studies were coded twice, and the intercoder agreement rate for the correlational outcomes was approximately 99%. Disagreements were resolved by consulting the original article or by discussion; whenever agreement was not reached, a third independent coder was consulted. The final list of studies included in the present meta-analysis is provided in the online supplement material; the list is organized in terms of overlap with the Swanson et al. (2003) sample. We also identified the sample and tasks used for each study (see the online supplemental material).

Recorded Variables and Coding

For each study, we recorded several variables that could potentially be important to explain the variation between studies (moderators). Several studies reported more than one measure of each of the target constructs (e.g., two similar tasks to assess word reading accuracy), and where this occurred, the individual effect sizes were aggregated using the arithmetic mean. This aggregation prior to meta-analytic integration is necessary to avoid over-representation of multi-experiment studies in the overall analyses (Rosenthal, 1991). The general categories of coding for each article included (a) the sample characteristics, (b) the orthography in which the study was conducted, and (c) the dependent measures included in the correlation matrix (RAN and reading).

Sample characteristics.

Grade. Grade range was coded to distinguish between different levels of reading acquisition. Samples consisting of kindergartners were coded as "pre-readers," samples containing readers in Grades 1 and 2 were coded as "beginning readers," samples with readers in Grades 3 and 4 were coded as "intermediate readers," and samples with a lower grade limit of Grade 5 or higher were assigned to the "advanced readers" category. When grade level was not reported, we estimated it on the basis of sample age (e.g., 6 years old = Grade 1; 7 years old = Grade 2). In cases of studies with samples covering grade ranges greater than 3 years and for which only the combined data were available, these samples were not included in the grade moderator test (21 independent effect sizes in total).

Reading group. The reading status of the sample was coded in three categories. Representative study samples, chosen using some form of random sampling for participant selection, were considered to be population based and were coded as "unselected sample." Samples including only dyslexic and/or poor readers were coded as "impaired readers," and study samples containing subjects described as having no reading problems were coded as "average readers." Whenever a study reported only the combined data from a sample with impaired readers and a sample with average readers, these effect sizes (17 in total) were excluded from the reading group moderator test.

Orthography.

Writing system. Each study was coded as "alphabetic" (e.g., English), "logographic" (e.g., Chinese), or "syllabic" (e.g., Japanese) writing system. The potential moderator effect of the writing system was evaluated by comparing the mean effect size estimated from studies in each of these categories. In the remainder of the meta-analysis, however, only the studies classified as "alphabetic" were included.

Consistency of orthography. The orthographic depth of the writing system in which the study was conducted was coded in three categories: "opaque" (e.g., English), "transparent" (e.g., Finish, Greek), and "intermediate" (e.g., Portuguese, Dutch) depth, following the classification system of Seymour et al. (2003).

RAN task.

RAN stimulus. RAN assessments were taken to include various measures of speeded naming of familiar items, presented in a continuous list format (multiple, matrix presentation of items) as in the classical RAN tasks (Denckla & Rudel, 1976). The tasks used to measure RAN performance were first categorized according to stimulus type. Stimulus type was coded to distinguish between

"letter naming," "digit naming," "picture naming," and "color naming." In several studies, continuous rapid naming of letters and numbers, or objects and colors, were combined to derive a single composite measure. In these cases, stimulus type was not coded and these effect sizes were excluded from the moderator analysis. Moreover, whenever Rapid Alternating Stimulus (RAS) tests (composed of alternating letters, numbers, and colors) were used, the corresponding effect sizes were not considered. Note that in the previous meta-analysis, Swanson et al. (2003) were more liberal in selecting studies based on the RAN task used, in the sense that they included studies that did not use the traditional RAN format.

RAN measurement. The method of assessment of RAN performance was also coded in two categories: "accuracy-based" assessment (e.g., proportion of items correctly named) and "fluency-based" assessment (e.g., items per second). However, because only a few independent effect sizes (less than five) were associated with a RAN measurement based on accuracy scores, these effect sizes were excluded from the analysis.

RAN format. We also recorded the total number of items to be named and the number of different token items included in the RAN task.

Reading skills.

Domain of reading assessed. The tests used to measure reading performance were first organized into four categories according to the domain of reading assessed: word reading, text reading, non-word reading, and reading comprehension. Word reading measures aim to assess visual recognition of real words by means of single-word identification tasks. To be termed a text reading, the task must involve reading sentences or text rather than single words. Non-word reading assesses word decoding skills; sample tasks include reading non-words or pseudo-words. Finally, reading comprehension consists of measures where a participant reads a passage or sentence and answers questions in relation to it; sample tasks include open-ended tests and multiple-choice-based tests. The reading measures were also coded so as to distinguish standardized measures of reading from non-standardized (experimental/experimenter developed) measures.

Type of reading score. The method of scoring reading performance was coded in two categories: accuracy-based assessment (e.g., proportion of correctly read words) and fluency-based assessment (e.g., words per second or correct words per minute, as in time-limited tests).

Meta-Analytic Procedures

Effect size estimates. The majority of the statistical analyses were conducted using the Comprehensive Meta-Analysis statistical software package (Borenstein, Hedges, Higgins, & Rothstein, 2005). The effect size index used in our meta-analysis was Pearson's correlation coefficient, r. By convention, an absolute effect size of .10 is considered small, a value of .30 is considered moderate, and .50 is considered a large effect (Cohen, 1988). We applied Fisher's r-to-z transformation to the effect size estimates to ensure the symmetry of the sampling distributions, as recommended in the literature (Borenstein, Hedges, Higgins, & Rothstein, 2009; Rosenthal, 1991). The z values were later transformed into r values for ease of interpretation.

As noted above, the studies in our sample often reported more than one effect size of interest for the same population. The validity of a meta-analysis rests on the assumption that each individual effect size is statistically independent of the others. There are several ways to incorporate multiple, dependent effect sizes into a meta-analysis while minimizing the likelihood of violating the independence assumption, but each of these methods involves tradeoffs. In our analysis, the shifting unit of analysis approach (Cooper, 2010) was employed, as this procedure provides a good compromise between preserving the independence of the effect sizes and retaining the maximum amount of information from each study. In this approach, each effect size associated with a sample is first coded as if it were an independent estimate of the relationship; the unit of analysis (samples, moderator modalities) is then shifted according to the hypothesis being tested. For the overall average effect analysis and whenever the moderator corresponds to a between-subjects factor defining separate groups of participants (grade, orthography, and reading group), we used the sample as unit of analysis; the multiple effects from each sample were aggregated, so each sample is only contributing one effect and all sample averaged effects are (almost) independent. When testing the moderator effect of an outcome domain (domain of reading assessed, scoring method of reading, and RAN stimulus type), and multiple effect sizes were available within the same sample, we shifted the unit of analysis from the sample to the effect sizes, allowing each sample to contribute one effect size to each category of the moderator. Admittedly, these effect sizes are not strictly independent; however, we decided not to restrict each sample to a single effect size to avoid an extensive loss of information. Yet, an additional sensitivity analysis was computed including only one randomly selected effect size per sample, in order to evaluate how ignoring the dependency between the effect sizes influenced the results. A similar methodological approach to handle this issue was already used in previous meta-analyses (e.g., Bramão, Reis, Petersson, & Faísca, 2011).

Analysis of effect sizes. After computing the effect sizes for each study, meta-analytic methods were applied to obtain a combined effect size indicating the magnitude of the association across all studies. The statistical significance of the mean effect size was tested with a random-effects model, which can account for systematic variation between studies (due to the moderators) and not only variation due to random error. Each effect size was weighted by multiplying its value by the inverse of its variance in order to correct for upward-biased estimates in small samples (Rosenthal, 1991). In addition, a 95% confidence interval (CI) was calculated; effect sizes are considered statistically significant when the CI does not include zero. A 95% prediction interval (PI) was also included (Borenstein et al., 2009). A variety of diagnostic tools were then used to search for outliers, including an analysis of the Cook's distance and studentized deleted residuals (Viechtbauer & Cheung, 2010). Whenever the presence of outliers was suspected, we performed a sensitivity analysis to examine their impact on the overall range of correlations. Thus, we repeated the meta-analysis multiple times, excluding successively all the potential outliers one by one up to the last and estimating the corresponding mean effect sizes.

To assess whether the variation in correlations between studies was significant, the homogeneity statistic Q was computed; a significant value on the Q test indicates a reliable variation between the correlations in the sample of studies. The I^2 statistic was used to estimate the percentage of variation across studies owing to

real heterogeneity rather than sampling error. For additional clarification of the differences between the effect size estimates, we conducted a subgroup analysis to test the moderator variable effects. A mixed effects model was employed to perform the moderator analysis, as this type of model allowed us to take into account some of the excess individual effect size variability that is not explained by the moderator being tested. Although conservative, these statistical models allow us to make inferences about the parameters of a population of studies (i.e., the universe of studies) that is larger than the set of observed studies and that may not be strictly identical to them (Hedges & Vevea, 1998). The degree of difference between the subsets of the studies was tested for significance by computing the $Q_{between}$ statistic, which has a chi-square distribution and is analogous to an analysis of variance (ANOVA) F test. In the case of the continuous moderator variables (total items and number of different tokens of the RAN task), a meta-regression based on the method of moments for random-effects model was used to predict the study outcome from the moderator variables. The statistical power of the moderator analyses was estimated using the procedures for mixed effects tests of moderators described by Hedges and Pigott (2004). These calculations were based on two-tailed inferential tests, the observed sample sizes and between-studies variance component, and a pre-specified difference between effect sizes. For this parameter we adopted a conservative approach and considered .07 as the smallest difference among group mean effect sizes that is of substantive importance; this value corresponds to a difference of .10 between Fisher's z values, which is, according to Cohen's (1992) convention, a small magnitude difference between two independent Pearson correlations.

Finally, to assess the possibility that publication biases affected our results, we performed a funnel plot analysis. In the absence of selection bias, this plot is expected to form an inverted funnel. The funnel is asymmetric in the presence of bias. The "trim and fill" method for random-effects models (Duval & Tweedie, 2000) was used to examine the impact of possible missing studies. This method imputes values in the funnel plot to render it symmetrical and calculates an estimated overall effect size on this basis. Because this method has been shown to spuriously adjust for nonexistent bias when the studies are heterogeneous (Terrin, Schmid, Lau, & Olkin, 2003), the publication bias analysis was complemented using a selection modeling approach proposed by Hedges (1992). This method assumes that the probability a result is selected for inclusion in a meta-analysis depends only on its p value, allowing to formally test the presence of publication bias with the likelihood ratio chi-squared statistic G^2 and to estimate an adjusted pooled effect. Computer code for selection modeling was made available by Vevea and Hedges (1995).

Results

The literature search yielded 126 articles containing 137 studies with samples from alphabetic writing systems, and 25 articles with samples from non-alphabetic writing systems, in which the relationship between RAN and reading performance was tested. The 137 studies reported 857 separate effect sizes based on 189 independent samples. In total, 28,826 subjects participated in these studies; individual effect sizes were reported for sample sizes ranging from 15 to 1,284.

Mean Effect Size Analyses

The first goal was to determine the average magnitude and significance of the relationship between RAN and reading performance. To do this, we first computed the mean effect size for the entire sample of studies and then examined each reading skill separately by calculating separate mean effect sizes for word reading, text reading, non-word reading, and reading comprehension (see Table 1).

Global mean effect size. An overall effect size was calculated incorporating all 189 independent effect sizes. The weighted mean r was .43, with a 95% CI of .41 to .45, 95% PI of .18 to .64. The homogeneity test was significant, $Q(188) = 594.86$, $p < .001$, $I^2 = 68.40$, and $\tau = 0.126$, suggesting that the variance in effect sizes could not be attributed to sampling error alone. A sensitivity analysis showed that after removing nine potential outliers, the overall correlation ranged between $r = .43$, 95% CI [.41, .45], and $r = .44$, 95% CI [.42, .47]. The funnel plot did not indicate any publication bias. This was confirmed by the selection modeling method ($G^2 = 9.01$, $df = 7$, $p = .251$). The results therefore indicated a moderate and significant relationship between RAN and reading performance, with substantial heterogeneity among the individual effect sizes.

Mean effect sizes for each domain of reading assessed.

Word reading. A total of 159 independent correlations, comprising 26,491 subjects, described the relationship between RAN and word reading. The weighted mean r was .45, with a 95% CI of .43 to .47, 95% PI of .16 to .67, $Q(158) = 618.12$, $p < .001$, $I^2 = 74.44$, and $\tau = 0.141$. After removing seven potential outliers using sensitivity analysis, the overall correlation range was $r = .45$, 95% CI [.43, .47] to $r = .46$, 95% CI [.44, .48]. The funnel plot indicated that studies were missing to the left of the mean. In a trim-and-fill analysis, five studies were added, and the adjusted overall correlation was $r = .44$, 95% CI [.42, .47]. The selection modeling method confirmed a publication bias ($G^2 = 13.69$, $df = 7$, $p = .057$).

Text reading. Text reading was associated with 27 independent effect sizes, comprising 2,798 subjects. The weighted mean r was .45, with a 95% CI of .38 to .50, 95% PI of .09 to .70, $Q(26) = 89.96$, $p < .001$, $I^2 = 71.10$, and $\tau = 0.159$. There were no outliers, and neither the funnel plot nor the selection modeling method ($G^2 = 7.22$, $df = 7$, $p = .406$) indicated any publication bias.

Non-word reading. One hundred thirteen independent effect sizes, comprising 17,019 subjects, were associated with non-word reading. The weighted mean r was .40, with a 95% CI of .38 to .42, 95% PI of .18 to .58, $Q(112) = 294.93$, $p < .001$, $I^2 = 62.03$, and $\tau = 0.106$. A sensitivity analysis showed that after removing five potential outliers, the overall correlation decreased to $r = .39$, 95% CI [.37, .41]. The funnel plot indicated that studies were missing to the right of the mean. In a trim-and-fill analysis, five studies were added, and the adjusted overall correlation was $r = .41$, 95% CI [.39, .43]. However, the selection modeling method did not corroborate that any publication bias existed ($G^2 = 5.84$, $df = 7$, $p = .558$).

Reading comprehension. Forty-three independent effect sizes, comprising 4,965 subjects, were associated with reading comprehension. The weighted mean r was .39, with a 95% CI of .34 to .44, 95% PI of .03 to .66, $Q(42) = 162.60$, $p < .001$, $I^2 = 74.17$,

Table 1

Overall Analysis and ANOVA Analog Test on the Effect Sizes Associated With the Four Domains of Reading Skills

Variable	N	k	n	Weighted mean r	95% CI	95% PI	τ	Q_{within}	$Q_{between}$
				Overall					
Total	137	189	28,826	.43	.41, .45	.18, .64	0.126	594.86***	
Trim-and-fill analysis									
Sensitivity analysis									
Min. (after excluding 1 PO)	—	—	—	.43	.41, .45	—	—	—	
Max. (after excluding 9 POs)	—	—	—	.44	.42, .47	—	—	—	
				Domain of reading					
Total									12.03**
Word reading	108	159	26,491	.45	.43, .47	.16, .67	0.141	618.12***	
Trim-and-fill analysis	—	—	—	.44	.42, .47	—	—	—	
Sensitivity analysis									
Min. (after excluding 1 PO)	—	—	—	.45	.43, .47	—	—	—	
Max. (after excluding 7 POs)	—	—	—	.46	.44, .48	—	—	—	
Text reading	23	27	27,98	.45	.38, .50	.09, .70	0.159	89.96***	
Trim-and-fill analysis									
Sensitivity analysis									
Non-word reading	78	113	17,019	.40	.38, .42	.18, .58	0.106	294.93***	
Trim-and-fill analysis	—	—	—	.41	.39, .43	—	—	—	
Sensitivity analysis									
Min./max.(after excluding 1–5 POs)	—	—	—	.39	.37, .41	—	—	—	
Reading comprehension	37	43	4,965	.39	.34, .44	.03, .66	0.162	162.60***	
Trim-and-fill analysis	—	—	—	.40	.35, .45	—	—	—	
Sensitivity analysis									
Min./max.(after excluding 1–2 POs)	—	—	—	.37	.32, .42	—	—	—	

Note. Dashes indicate not reported values, to avoid overloading the table. ANOVA = analysis of variance; N = number of studies; k = number of effect sizes; n = total sample size; CI = confidence interval; PI = prediction interval; Q_{within} = within-group homogeneity of variance; $Q_{between}$ = between-groups homogeneity of variance; min. = minimum; max. = maximum; PO = potential outlier.
** $p < .01$.   *** $p < .001$.

and $\tau = 0.162$. A sensitivity analysis showed that, after removing two potential outliers, the overall correlation decreased to $r = .37$, 95% CI [.32, .42]. The funnel plot indicated that studies were missing to the right of the mean. In a trim-and-fill analysis, two studies were added and the adjusted correlation was $r = .40$, 95% CI [.35, .45]. However, the selection modeling method did not corroborate that any publication bias existed ($G^2 = 5.52$, $df = 7$, $p = .597$).

These results indicate that RAN is significantly related to reading performance, regardless of the domain of reading skill assessed. Yet, word reading and text reading show the highest mean correlations with RAN ($r = .45$). Because there was some heterogeneity among the individual effect sizes included in the global mean effect size analysis, we next performed an ANOVA analog analysis on the effect sizes associated with the four domains of reading skills, to examine whether the domain of reading assessed accounts for some of this heterogeneity (see Table 1). ANOVA analog analyses group independent effect sizes into mutually exclusive categories on the basis of an independent variable (e.g., domain of reading skill) and resolve the total homogeneity statistic Q into the portion explained by the categorical variable ($Q_{between}$) and the residual pooled-within-groups portion (Q_{within}; Lipsey & Wilson, 2001). A significant heterogeneity across reading domains was observed, $Q_{between}(3) = 12.03$ and $p = .007$, suggesting therefore that the domain of reading assessed may to some extent explain part of the heterogeneity among the effect sizes included in the global mean effect size analysis. A significant result was also observed when we performed the sensitivity analysis, $Q_{between}(3) = 7.94$ and $p = .047$, using exclusively independent effect sizes. The

results of pairwise contrasts indicated that significantly higher coefficients emerged in favor of word reading measures when compared with non-word reading, $Q_{between}(1) = 8.85$, $p = .003$, and also with reading comprehension, $Q_{between}(1) = 5.72$, $p = .017$. No significant differences were found between word reading and text reading ($p = .828$, power = .73).

The residual heterogeneity was significant, $Q_{within}(338) = 1,165.61$ and $p < .001$, implying that unexplained variability remained. Further meta-analytic subdivision of the overall sample is therefore warranted.

Moderator Analyses

Additional analyses examined the role of theoretically meaningful moderator variables selected a priori to test for systematic differences between studies that could potentially explain the observed heterogeneity among the effect sizes (see Table 2). These moderator analyses were performed only on the effect sizes associated with word reading. We took this decision for two main reasons. First, because we found a significant effect of domain of reading assessed on the RAN–reading relationship, we could not be sure there was no difference in the moderator effects across measures. Thus, to avoid potential confounding effects we decided not to aggregate all the reading domains in the moderator analysis. Second, there might be other potential confounds when collapsing the reading measures, as for some measures there are categories within the moderators that are missing. For example, for reading comprehension almost all of the individual effect sizes were asso-

Table 2

Moderator Analysis and ANOVA Analog Test of Effect Size Homogeneity for Moderators of the Relationship Between Rapid Naming and Real Word Reading

Moderator	N	k	n	Weighted mean r	95% CI	95% PI	τ	$Q_{between}$
Unit of analysis: Effect sizes								
Reading outcome								9.90**
Reading accuracy	79	101	12,239	.42	.39, .45	.11, .66	0.153	
Reading fluency	55	92	15,710	.49	.46, .52	.19, .71	0.149	
Opaque orthographies only								13.14***
Reading accuracy	69	83	8,913	.44	.41, .49	.16, .66	0.139	
Reading fluency	28	33	6,565	.55	.51, .60	.15, .78	0.187	
Transparent orthographies only								8.62**
Reading accuracy	10	15	3,006	.35	.28, .41	.07, .58	0.143	
Reading fluency	25	50	8,263	.45	.42, .48	.22, .63	0.097	
RAN stimulus type								82.74***
Letters	55	82	13,124	$.51_a$	.48, .53	.25, .70	0.136	
Numbers	60	95	12,622	$.48_a$	.46, .50	.23, .66	0.120	
Pictures	32	56	8,409	$.35_b$	.32, .38	.16, .50	0.082	
Colors	25	33	2,402	$.33_b$	.27, .38	.27, .38	0.014	
Unit of analysis: Sample								
Grade								
Reading accuracy only								9.30*
Prereaders	9	13	1,902	$.41_a$	.33, .49	.05, .68	0.116	
Beginner readers	24	29	3,971	$.47_a$	.42, .52	.16, .70	0.180	
Intermediate readers	16	20	1,743	$.35_b$	.27, .42	.01, .62	0.062	
Advanced readers	10	11	2,027	$.35_b$	.25, .44	.05, .65	0.132	
Reading fluency only								3.28
Prereaders	5	5	1,886	.40	.26, .52	.25, .80	0.149	
Beginner readers	25	37	6,765	.45	.40, .49	.14, .67	0.163	
Intermediate readers	11	24	3,333	.50	.44, .55	.19, .71	0.086	
Advanced readers	6	8	1,453	.48	.39, .57	.07, .75	0.126	
Opaque orthographies only								10.28*
Prereaders	10	10	2,753	$.43_a$	.34, .50	.06, .69	0.144	
Beginner readers	25	28	4,818	$.50_{a,b}$	.45, .55	.21, .71	0.139	
Intermediate readers	15	18	1,145	$.46_{a,b}$	.36, .56	.05, .74	0.143	
Advanced readers	7	8	972	$.36_{a,c}$	.28, .43	.02, .62	0.090	
Transparent orthographies only								4.37
Prereaders	4	7	1,035	.38	.28, .48	.03, .68	0.071	
Beginner readers	16	22	3,137	.43	.38, .48	.15, .64	0.128	
Intermediate readers	10	19	3,288	.47	.41, .52	.19, .68	0.093	
Advanced readers	6	8	1,809	.37	.28, .46	.01, .66	0.180	
Reading group								3.37
Average readers	20	24	2,025	.45	.39, .51	.14, .68	0.186	
Impaired readers	24	27	2,152	.49	.44, .54	.20, .70	0.102	
Unselected sample	58	98	18,415	.43	.41, .46	.15, .65	0.134	
Consistency of orthography								
Alphan. RAN and reading fluency								13.90***
Opaque	18	21	4,895	$.57_a$	.53, .61	.38, .72	0.097	
Transparent	20	45	6,764	$.48_b$	.45, .51	.27, .64	0.108	
Intermediate	3	8	701	$.54_{a,b}$	.45, .62	.25, .74	0.044	
Reading accuracy only								6.87*
Opaque	59	74	8,332	.44	.41, .48	.14, .67	0.143	
Transparent	10	15	3,006	.35	.27, .42	.01, .63	0.143	
Writing system								
Reading fluency only								3.87*
Alphabetic	34	74	12,866	.48	.45, .51	.18, .69	0.152	
Non-alphabetic	16	21	3,077	.54	.49, .60	.24, .75	0.094	
Reading accuracy only								.01
Alphabetic	54	71	9,594	.42	.39, .46	.04, .68	.174	
Non-alphabetic	25	37	3,555	.42	.37, .47	.29, .54	.063	

Note. Different subscripts indicate weighted mean effect sizes that are significantly different. ANOVA = analysis of variance; N = number of studies; k = number of effect sizes; n = total sample size; CI = confidence interval; PI = prediction interval; $Q_{between}$ = between-groups homogeneity of variance; RAN = rapid automatized naming; Alphan. = alphanumeric.
* $p < .05$. ** $p < .01$. *** $p < .001$.

ciated with accuracy-based reading scores, only one effect size was associated with samples of advanced readers, and no effect size was associated with samples coming from orthographies with intermediate depth. Thus, the moderator analysis was restricted to word reading measures because these are significantly heterogeneous and considerably larger in number, and therefore more variability exists among the individual effect sizes linked to the moderators of interest. The categorical moderators were tested using ANOVA analog analyses. Pairwise contrasts were tested for the moderators involving more than two modalities. In these cases, we applied the Bonferroni correction method with a significance threshold of .05/L, where L is the number of contrasts to be conducted. When fewer than five studies reported data on a moderator, that variable was excluded from the analysis. Additional analyses were conducted to examine the impact of continuous variables (RAN format: total number of items, and number of different tokens) on effect size using meta-regression.

We should note that we found no differences between studies using reading standardized assessments and those with experimenter-developed measures, $Q_{\text{between}}(1) = 0.04$ and $p = .850$. Then, this variable was not considered thereafter.

Type of reading score. We tested whether the strength of the relationship of RAN and reading was dependent on the way reading performance is assessed. Across languages, the comparison between studies that used fluency-based assessments and studies that used accuracy-based assessments yielded significant results, with the former group displaying higher effect sizes. Within each orthography, we examined whether the correlations obtained between the RAN task and the reading accuracy measure differed significantly from the correlations obtained between the RAN task and the reading fluency. The results demonstrated that the difference between the correlations was significant in all orthographies, with RAN being more strongly related to reading fluency than to reading accuracy. A similar results pattern emerged when a sensitivity analysis using solely independent effect sizes was performed.

RAN stimulus. We predicted that the magnitude of the correlation between RAN and reading performance would be higher for alphanumeric stimulus items (e.g., letters, numbers), especially letters, compared to non-alphanumeric stimulus items (e.g., colors, pictures). This hypothesis was supported by the ANOVA analog analyses, showing that RAN stimulus type was a significant moderator variable. This result was confirmed when we carried out a sensitivity analysis using exclusively independent effect sizes. Studies that used RAN tasks with letters or numbers showed higher correlations with reading than did studies that used RAN tasks with colors or pictures (all comparisons with $p < .001$). The results of the pairwise contrasts also indicated that the correlations with letter naming tended to be somewhat higher than those with number naming, although not significantly so, $Q_{\text{between}}(1) = 1.72$, $p = .19$; power $= .98$. Moreover, the mean effect size associated with picture naming was not significantly larger than the mean effect size associated with color naming, $Q_{\text{between}}(1) = .62$, $p = .43$; power $= .92$.

RAN format. Another goal in our meta-analysis was to evaluate whether the number of different items to be named (i.e., the number of different tokens) and the total number of items had a significant impact on the effect size. Meta-regression analyses showed that neither the number of tokens ($\beta = -.003$, $p = .76$, $k = $ 135, $R^2 = .019$) nor the total number of RAN items ($\beta = -.000$, $p = .63$, $k = 148$, $R^2 = .007$) had a significant impact on the correlation magnitude.

Grade. One of our goals was to examine the extent to which the between-studies variability in grade accounted for the inconsistencies in the magnitude of the correlations between RAN and reading performance. When RAN stimulus types and reading outcomes were all considered, grade had no significant impact on the correlation magnitude, $Q(3) = 3.12$, $p = .37$. Consistent results were obtained only for studies measuring RAN and reading fluency. For correlations between RAN and reading accuracy, the ANOVA analog analysis revealed significant heterogeneity across grade groups. In particular, the data indicated that the effect sizes follow a non-linear trajectory; the effect sizes increased, although not significantly, $Q_{\text{between}}(1) = 1.12$, $p = .29$; power $= .35$, from kindergarten to first/second grade, and after that, they decreased. The mean effect size was smaller for intermediate readers than for beginning readers, $Q_{\text{between}}(1) = 6.37$, $p = .01$, and was stable thereafter, as indicated by a non-significant contrast between intermediate readers and advanced readers, $Q_{\text{between}}(1) = 0.06$, $p = .81$, possibly owing to the low level of statistical power in this analysis (power $= .41$).

We also sought to evaluate whether the impact of grade was different across different orthographies. When RAN stimulus types and type of reading score were all considered, grade had a significant impact on the effect size for opaque orthographies, $Q(3) = 10.28$, $p = .02$. The effect of RAN was greatest in kindergarten and first/second grade and declined thereafter. For transparent orthographies, on the other hand, grade did not have a significant impact on the correlation magnitude, $Q(3) = 4.37$, $p = .225$, although these results should be viewed with caution due to the lack of power (power $= .26$). Unfortunately, the number of studies was too small to re-analyze the data with the effects of RAN stimulus type and reading outcome controlled for simultaneously.

Reading group. To examine the moderator effect of reading group, we compared studies with a representative sample (unselected) to those sampling average or impaired readers. As shown in Table 1, the mean effect size was larger for studies using impaired reader samples than for studies with average readers or unselected samples; however, these differences were not statistically significant, perhaps owing to the low level of statistical power in this analysis (power $= .65$). Similar non-significant differences were found after splitting our sample of studies into two halves based on the type of reading score (accuracy- vs. fluency-based) and when we considered only the effect sizes obtained from studies in which alphanumeric RAN was measured.

Consistency of orthography. To study the moderator role of orthographic consistency, we first compared the correlations between RAN and the reading fluency measure across languages, and then compared the correlations between RAN and the reading accuracy measure across languages. The relationship between RAN alphanumeric tasks and reading fluency was significantly stronger for opaque orthographies than it was for transparent orthographies, $Q_{\text{between}}(1) = 13.10$, $p < .001$. The mean effect size associated with intermediate orthographies was not significantly different from either the mean effect size associated with opaque orthographies, $Q_{\text{between}}(1) = .64$, $p = .42$; power $= .35$, or the mean effect size associated with transparent orthographies, $Q_{\text{between}}(1) = 1.75$, $p = .19$; power $= .35$. For associations with

reading accuracy, RAN stimulus types were all considered because the number of independent effect sizes associated with alphanumeric RAN was not sufficient to isolate the relationship between alphanumeric RAN and reading accuracy. In this analysis, the mean effect size associated with opaque orthographies was again significantly larger than the mean effect size associated with transparent orthographies, $Q_{between}(1) = 5.05$, $p = .02$.

Writing system. The moderator effect of the writing system was examined by comparing studies that were conducted in an alphabetic writing system with those that were conducted in non-alphabetic writing systems (logographic or syllabic). No significant difference was found, $Q_{between}(1) = 0.33$, $p = .56$; power = .90; both types of studies displayed a significant and moderate RAN–reading relationship. However, in the second ANOVA analog, when the analysis was performed separately according to type of reading score, we found that the correlations of RAN with reading fluency were significantly higher for non-alphabetic writing systems than for alphabetic writing systems (.54 and .48, respectively).

Discussion

Evidence that rapid serial naming skills are associated with reading ability has become increasingly prevalent in recent years (for a review, see Kirby et al., 2008; Norton & Wolf, 2011). However, there are large variations between studies in the magnitude of the reported RAN–reading correlations. The extent to which differences in measures and samples across studies may act as moderators of this association is as yet poorly understood. Our aim in this meta-analysis was to determine the mean magnitude of the correlation between RAN and reading performance and to explore factors that could potentially account for the heterogeneity between studies.

The overall meta-analysis unambiguously demonstrated that there is a significant moderate-to-strong correlation ($r = .43$) between RAN performance and reading ability, which disagrees with some earlier reports (e.g., Scarborough, 1998). The global mean effect size is also similar to the correlation of .46 obtained in a previous meta-analysis (Swanson et al., 2003). Thus, our meta-analysis demonstrated that RAN tasks have great potential in predicting reading ability, presumably because performance in both RAN and reading taps into shared cognitive processes. The results of the overall analysis also showed a substantial degree of heterogeneity among the individual effect sizes ($I^2 = 68.4$) and thus suggest that the variability in the RAN–reading association observed in the literature is likely related to certain specifics of the studies. Consistent with this, five of the moderators that we tested accounted for a significant part of the between-effect-sizes variability observed.

An important question is whether the RAN construct is associated with specific domains of reading. In this study, we found RAN to be linked not only to the domain of reading most frequently assessed in the field (real word reading) but also with text reading, non-word reading, and reading comprehension. This finding implies that rapid naming correlates with reading performance regardless of whether the literacy measure relies more heavily on phonological or orthographic coding skills. Even though the overlapping variance between RAN and reading that is shared among the reading outcomes cannot be estimated from bivariate correla-

tional evidence, it is likely that variations in RAN performance might reflect a common mechanism underlying these four distinct measures of reading. Nevertheless, the size of the mean correlation with RAN appears to depend on the domain of reading assessed, as significantly higher coefficients emerged in favor of real word reading and text reading. However, based on univariate analysis alone, one cannot rule out the possibility that statistical differences across these reading domains are due to uncontrolled moderators that might differ across studies, and, therefore, conclusions on these differences should be regarded cautiously.

In interpreting the mechanisms behind these findings, neither a unitary orthographic explanation (e.g., Bowers & Wolf, 1993) nor a phonological interpretation (e.g., Torgesen et al., 1997) for the relation between RAN and reading are fully supported. We hypothesized that if RAN were measuring a cognitive mechanism relevant for orthographic processing (Bowers & Newby-Clark, 2002; Bowers & Wolf, 1993), then it should be more important for word and text reading than for non-word reading, particularly in the domain of fluency, because the former involve a higher amount of orthographic processing. Although our results were in the expected direction, with RAN being a better correlate of word reading and text reading ($r = .45$), the fact that it also correlates moderately with non-word reading ($r = .40$), which is a classical measure of phonological decoding, may still hold for a phonological processing role. It should be noted, however, that the non-words results can also be explained in an orthographic context if we assume that orthographic knowledge includes orthographic information at the sub-word level (Bowers & Wolf, 1993; however, see Georgiou et al., 2009). In most studies, non-words and pseudo-words were derived from real words and therefore have familiar letter patterns that can be analyzed as orthographic units. Thus, most measures of reading ability actually depend on both phonological and orthographic processes, or at least allow readers to use both, and therefore may not be the most suitable in addressing the aforementioned hypotheses.

We also distinguished between two ways of assessing of reading performance: accuracy-based and fluency-based. The results clearly demonstrated that the magnitude of the correlation between the RAN and reading measures depends on the use of accuracy versus fluency of word recognition as the outcome variable. As we expected, RAN was strongly related to reading fluency ($r = .49$ vs. $r = .42$ for accuracy). This result is not surprising given the overlap between RAN and reading fluency, with both measures sharing a time component. Some authors have claimed that RAN and reading are linked because skilled performance in both depends on the rapid execution of the underlying processes (Kail & Hall, 1994; Kail, Hall, & Caskey, 1999). However, invoking processing speed as the single explanation for the correlation between RAN and reading is insufficient, because in this case a significant relationship with reading accuracy would not have been apparent. Other studies have also disfavored a general processing speed explanation (see, e.g., Vaessen, Gerretsen, & Blomert, 2009).

Another hypothesis we tested was whether the association between RAN and reading in transparent orthographies was to a strong degree moderated by the type of reading score when compared with opaque orthographies. This is to be expected because individual differences in the speed of word reading tend to be particularly important among children learning to read in a trans-

parent orthography, and accuracy rapidly converges on an asymptote (Wimmer, 2006). In this meta-analysis, the relationship patterns between RAN performance and scoring method of reading were found to be similar across orthographies. Unfortunately, there were too few studies to analyze whether these findings remain unchanged after accounting for variations in the samples as a function of grade.

Several modalities and formats have also been used to assess RAN performance. Here, we explored the moderator effects of two methodological features of RAN assessments: the use of alphanumeric versus non-alphanumeric items, and the total number of items and number of different tokens in the RAN task. In agreement with previous studies (Lervåg & Hulme, 2009; van den bos et al., 2003), the performance measured on letter- and digit-naming tasks was more strongly related to reading competence than on tasks requiring naming of colors or objects. Thus, alphanumeric RAN tasks seem to better capture underlying processing abilities that are important for word reading and therefore should be preferred over non-alphanumeric ones whenever a prediction of the reading ability is of interest. One question that follows is whether rapid naming of letters is just a simple version of a reading fluency test. Bowey, McGuigan, and Ruschena (2005) argued that the correlation of alphanumeric RAN tasks with reading ability may simply be an effect of the letter–sound knowledge of the children; that is, weak letter knowledge will lead to both slow RAN for letters and problems in learning to read. However, we argue that it is unlikely that letter knowledge per se is the sole reason why RAN is associated with reading. If the predictive power of RAN tasks were dependent on letter name knowledge, then picture- and color-naming speed should not have been correlated with reading outcomes.

In addition to testing the moderator effect of RAN stimulus type, we tested a new hypothesis concerning the impact of the resource demands of the RAN tasks, parameterized here by the total number of items and the number of different tokens. For example, the version used in Cobbold and colleagues' (2003) study comprised 20 pictured objects with no repetitions. Arguably, a test like this could place greater demands on a participant's ability to recall items from lexical memory than would the traditional RAN task (in which five items are repeated 10 times; Norton & Wolf, 2011). However, our meta-analytic results showed no significant moderator effect, even if the amount of cognitive resources used in the encoding and retrieval processes may increase with the total number of RAN items and the number of different tokens. In sum, RAN tasks with letters or numbers (vs. non-alphanumeric items) correlate strongly with current reading, but the strength of the cognitive contribution of RAN to reading performance keeps preserved even when the classical RAN paradigm is modified in terms of the resource demands.

Some researchers have also examined developmental differences in the RAN–reading relationship. Based on developmental theories of reading (e.g., Ehri, 2007), as the reading competence progresses the relative involvement of cognitive processes underlying reading would be expected to change. Accordingly, some studies have suggested that the magnitude of the RAN–reading association may change over the course of reading development, although the direction of the change is as yet undetermined. Some authors have suggested that the impact of RAN on reading performance increases as children become more proficient readers

(e.g., Vaessen et al., 2010; Vaessen & Blomert, 2010), whereas others hold that the role of RAN is time-limited (e.g., Torgesen et al., 1997). First, our results indicated that there is an association between rapid naming skill and reading performance from the very beginning of reading acquisition, which is consistent with previous reports (e.g., Compton, 2003; Lervåg, Bråten, & Hulme, 2009). Moreover, the influence of RAN on reading continues throughout elementary school, as we found a significant correlation across all grades sampled. A previous meta-analysis (Swanson et al., 2003) found that correlations between reading and RAN varied minimally across age. In our meta-analysis we extended this result by showing that the RAN association evolves differently over time for different reading scoring methods, as already suggested by others (Georgiou, Parrila, Kirby, et al., 2008; Torgesen et al., 1997). Specifically, our results demonstrated that the RAN–reading fluency coefficients were stable across grades, as in a longitudinal study from Grades 1 to 8 (Landerl & Wimmer, 2008), but relations with reading accuracy declined with increasing reading experience (i.e., are moderated by grade level): The effect sizes associated with reading accuracy increased from kindergarten to first/second grade (i.e., beginning readers) and decreased thereafter as the subjects become more proficient readers. It might be the case that our findings result from a methodological artifact due to ceiling effects at higher grades for the accuracy measures. Alternatively, the attenuation pattern in accuracy may result from a true reduction of individual differences in reading accuracy as children get older (i.e., the underlying ability may itself attain a ceiling). In both cases, the resulting homogeneity would lead to weaker correlations between RAN and reading accuracy, while no asymptote is expected for reading fluency. For example, in Høien-Tengesdal's (2010) results the performance of the sixth-grade children on word identification was close to ceiling (accuracy: 94.2%), which perhaps explains the small correlation that was observed with RAN performance ($r = .13$). Similar ceiling effects for accuracy measures were also identified at least in two more studies (Georgiou, Parrila, & Liao, 2008; Kairaluoma, Torppa, Westerholm, Ahonen, & Aro, 2013). However, for the majority of available data there were no notable ceiling effects for word reading accuracy in intermediate and advanced grades (at least 12 out of 15 studies using no time-limited tasks), which might suggest a real attenuation for the RAN–reading accuracy association throughout development. We should note that 11 out of these 12 studies were conducted in opaque orthographies, and so the results must be interpreted with caution, as they may not generalize to transparent orthographies. In summary, the data suggest that RAN taps a key dimension of reading ability with a persistent effect over the course of development, even though its relationship with various types of reading measurements (accuracy and fluency) is somewhat different across grade levels. As recently raised by Protopapas, Altani, and Georgiou (2013), it is possible that RAN and reading may involve one common set of cognitive mechanisms in early development, contributing to both reading accuracy and speed, and a different common set of cognitive mechanisms after expertise is attained, less important for accuracy. Clearly, further investigation is warranted for a better understanding of the developmental interrelations among RAN and reading tasks.

In reviewing the RAN–reading research, we also found that children's reading competence (average vs. impaired readers) could, to some degree, moderate the correlations between RAN

and reading tasks. Studies that have directly compared samples of average readers to samples of impaired readers tend to report a stronger RAN–reading relationship for the latter samples (e.g., Araújo et al., 2011). This meta-analysis only partially supports this finding; although the results were in the expected direction, with the mean effect size for impaired readers ($r = .49$) being greater than for average readers ($r = .45$) and unselected samples ($r = .43$), the moderator effect of group was not statistically significant (even after controlling for the RAN stimulus type and reading score). In their previous meta-analyses, Swanson and colleagues (2003) also found small differences between poor and skilled readers in RAN and reading correlation ($r = .41$ and $r = .43$, respectively). It should be noted, however, that owing to the low level of statistical power in our analysis, we cannot discard the possibility that reading group does account for the heterogeneity in the overall effect sizes distribution. Still, the fact that we had observed a significant and large correlation in the impaired readers sample suggests that the cognitive processes required by RAN tasks are important factors in reading difficulties and that these tasks may therefore constitute a useful tool for predicting reading failure. The reason why rapid naming skills would play a more important role among impaired as opposed to average readers is not completely understood. One possible interpretation was discussed by McBride-Chang and Manis (1996), who argued that good readers are nearly "automatic" in their performance in the early stages of reading development, while impaired readers rarely approach levels of automaticity. A task such as RAN is more difficult (Araújo et al., 2010; Fawcett & Nicolson, 1994) and is more likely to create variability in impaired readers; consequently, RAN might be a better correlate of reading skill in these readers.

Finally, we also examined whether the depth of the orthography impacts the correlations between RAN and reading performance. Previous studies have obtained inconsistent results on this topic (e.g., Georgiou, Parrila, & Papadopoulos, 2008; Patel et al., 2004). Our data demonstrated that RAN is a significant correlate of reading ability across languages that vary in orthographic consistency, and also that the orthographic depth moderates the magnitude of these correlations for both reading fluency and reading accuracy. Contrary to our expectations, however, the correlations were weaker for transparent than for opaque orthographies. Some authors have claimed an inverse pattern, in other words, that RAN plays a greater role in transparent orthographies (de Jong & van der Leij, 2003; Wimmer, Mayringer, & Landerl, 2000). A possible explanation is the preferential use of reading fluency measures in published studies with transparent orthographies, which may have accentuated the role of RAN, creating an appearance of greater involvement of RAN in these orthographies. Yet, the results of our study contradict those of two previous cross-linguistic studies using comparable reading measures: Both Patel and colleagues (2004) and Ziegler and colleagues (2010) observed that RAN is equally important in opaque and transparent orthographies. These discrepancies likely reflect the fact that these two cross-linguistic studies have only used non-alphanumeric RAN measures. This may have underestimated the RAN–reading relationship, since alphanumeric RAN tasks are more closely associated with reading ability than are non-alphanumeric RAN, and so differences across languages were not captured. Another interesting outcome from the present study was that the RAN–reading relationship seems to follow different developmental patterns in opaque and transparent

orthographies. For opaque orthographies, the correlations were moderated by the grade level of the children, with somewhat larger effects in early grades, while for transparent orthographies they seemed to vary minimally across grade levels (i.e., were developmentally stable). In sum, the meta-analytic results converge on the conclusion that RAN reflects a critical aspect of the "cognitive foundations" for learning to read.

In addition, cross-orthographies exploration in this study reveals that RAN plays an important role in learning to read regardless of whether the sample subjects used an alphabetic or non-alphabetic writing system, such as Chinese or Japanese. When reading fluency measures were considered, RAN is an even stronger correlate of reading ability in non-alphabetic than in alphabetic languages ($r = .54$ and $r = .48$, respectively). This finding is in agreement with the only cross-linguistic study that examined RAN and reading across different writing systems (English-, Greek-, and Chinese-speaking children; Georgiou, Parrila, & Liao, 2008). Thus, the data appear to suggest that reading fluency in non-alphabetic languages imposes greater demands on shared underlying rapid naming processes, although it is not clear which relevant cognitive constructs are captured by RAN in these languages. The need for identifying visual-orthographic information is central in Chinese (Shen & Bear, 2000; Shu, 2003), and because RAN tasks also require children to identify items by incorporating visual-orthographic information, it is tempting to interpret this factor as the basis of the (stronger) association between RAN and reading ability.

Limitations

The current meta-analysis has some limitations. First, our meta-analysis was limited to published studies, and the outcomes therefore may reflect publication biases. Significant results are more likely to be published than null results, and so the true mean effect size may be overestimated (Begg, 1994). However, we tried to overcome this limitation by using statistical procedures to detect and estimate the presence of publication bias.

Besides, insufficient data occasionally prevented examination of higher order interactions among moderators, because categories did not yield enough coefficients and there are gaps in literature—for example, our database contained too few correlations involving orthographies of intermediate depth, and this can have relevance for the understanding of cross-language differences in reading. Another important issue is the lack of statistical power of some of our analyses, due to the fact that the modalities of some moderators were reported in only a small number of studies. Therefore, it is not possible to make strong claims, for example, about reading level effects (average vs. impaired readers) on the magnitude of the correlations. Hence, this analysis should be viewed as suggestive rather than conclusive. Admittedly, also, the separate effect size estimates in the same study are not always completely independent—for example, reading groups in the same study might be broken out into different samples (e.g., age groups) and their data analyzed separately. Although these groups share methodological and situational influences (e.g., the same laboratory, the same researcher, the same data collection procedures) and consequently are not strictly independent, we adopted the shifting unit of analysis approach (Cooper, 2010) that is not too conservative and maximizes the information available. Last, we should note that all

the data included in the meta-analysis are cross-sectional and therefore cannot inform about the direction of causality. While we provide evidence that RAN is an important correlate of reading skills, we cannot distinguish whether there is a reciprocal influence of reading fluency/accuracy on the growth of RAN skill.

Conclusions and Recommendations for Future Studies

This meta-analysis provides convincing evidence for a significant and persistent relation between RAN and reading ability. It is suggested that RAN performance reflects, from very early on, underlying cognitive processes that are relevant for learning to read and, consequently, reflects the usefulness of these measures in predicting reading competence and its failure. Correlations were higher for reading fluency than for accuracy measures, and also when alphanumeric RAN stimulus material was used. Thus, these measures should be critical in predictive studies, especially at more advanced school grades, as they seem to be more sensitive to individual differences in reading ability than are other measures. But whatever RAN taps into, it is beyond letter knowledge and speed of processing. In turn, the association of RAN with reading accuracy seems to depend greatly on the grade level of the sample under study, with somewhat larger effects in early grades. Later on in development, we recommend that researchers ensure the discriminative power of the accuracy measures, due to possible ceiling effects. Moreover, we also observed that correlations tend to be stronger for impaired readers than for average readers. Previous studies had already shown that children with dyslexia perform poorly on RAN (e.g., Araújo et al., 2011, 2010). Taken together, we argue that understanding the processes subserving impaired visual naming speed in dyslexia may offer promising clues to the causes of dyslexia and deepen our knowledge of the cognitive basis of this disorder. Finally, our cross-linguistic comparisons suggest that processes underlying RAN are universal influence on reading system, although their relative importance is stronger for opaque, in particular for non-alphabetic, orthographies.

References

Araújo, S., Inácio, F., Francisco, A., Faísca, L., Petersson, K. M., & Reis, A. (2011). Component processes subserving rapid automatized naming in dyslexic and non-dyslexic readers. *Dyslexia, 17,* 242–255. doi:10.1002/dys.433

Araújo, S., Pacheco, A., Faísca, L., Petersson, K. M., & Reis, A. (2010). Visual rapid naming and phonological abilities: Different subtypes in dyslexic children. *International Journal of Psychology, 45,* 443–452. doi:10.1080/00207594.2010.499949

Babayiğit, S., & Stainthorp, R. (2010). Component processes of early reading, spelling, and narrative writing skills in Turkish: A longitudinal study. *Reading and Writing, 23,* 539–568. doi:10.1007/s11145-009-9173-y

Begg, C. B. (1994). Publication bias. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 399–410). New York, NY: Russell Sage Foundation.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive Meta-Analysis Version 2* [Computer software]. Engelwood, NJ: Biostat.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. Chichester, United Kingdom: Wiley. doi:10.1002/9780470743386

Bowers, P. G., & Newby-Clark, E. (2002). The role of naming speed within a model of reading acquisition. *Reading and Writing, 15,* 109–126. doi:10.1023/A:1013820421199

Bowers, P. G., & Wolf, M. (1993). Theoretical links among naming speed, precise timing mechanisms and orthographic skill in dyslexia. *Reading and Writing, 5,* 69–85. doi:10.1007/BF01026919

Bramão, I., Reis, A., Petersson, K. M., & Faísca, L. (2011). The role of color information on object recognition: A review and meta-analysis. *Acta Psychologica, 138,* 244–253.

Caravolas, M., Volin, J., & Hulme, C. (2005). Phoneme awareness is a key component of alphabetic literacy skills in consistent and inconsistent orthographies: Evidence from Czech and English children. *Journal of Experimental Child Psychology, 92,* 107–139. doi:10.1016/j.jecp.2005.04.003

Chiappe, P., Stringer, R., Siegel, L. S., & Stanovich, K. E. (2002). Why the timing deficit hypothesis does not explain reading disability in adults. *Reading and Writing, 15,* 73–107. doi:10.1023/A:1013868304361

Clarke, P., Hulme, C., & Snowling, M. (2005). Individual differences in RAN and reading: A response timing analysis. *Journal of Research in Reading, 28,* 73–86. doi:10.1111/j.1467-9817.2005.00255.x

Cobbold, S., Passenger, T., & Terrel, C. (2003). Serial naming speed and the component elements of speech time and pause time: Relationships with the development of word-level reading in children aged four to five years. *Journal of Research in Reading, 26,* 165–176. doi:10.1111/1467-9817.00194

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159. doi:10.1037/0033-2909.112.1.155

Compton, D. L. (2003). Modeling the relationship between growth in rapid naming speed and growth in decoding skill in first-grade children. *Journal of Educational Psychology, 95,* 225–239. doi:10.1037/0022-0663.95.2.225

Cooper, H. M. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Thousand Oaks, CA: Sage.

Cunningham, A. E. (2006). Accounting for children's orthographic learning while reading text: Do children self-teach? *Journal of Experimental Child Psychology, 95,* 56–77.

Cutting, L. E., & Denckla, M. B. (2001). The relationship of serial rapid naming and word reading in normally developing readers: An exploratory model. *Reading and Writing, 14,* 673–705. doi:10.1023/A:1012047622541

de Jong, P. F. (2011). What discrete and serial rapid automatized naming can reveal about reading. *Scientific Studies of Reading, 15,* 314–337. doi:10.1080/10888438.2010.485624

de Jong, P. F., & van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology, 91,* 450–476. doi:10.1037/0022-0663.91.3.450

de Jong, P. F., & van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology, 95,* 22–40. doi:10.1037/0022-0663.95.1.22

Denckla, M. B., & Cutting, L. E. (1999). History and significance of rapid automatized naming. *Annals of Dyslexia, 49,* 29–42. doi:10.1007/s11881-999-0018-9

Denckla, M. B., & Rudel, R. G. (1976). Rapid "automatized" naming (R. A. N.): Dyslexia differentiated from other learning disabilities. *Neuropsychologia, 14,* 471–479. doi:10.1016/0028-3932(76)90075-0

Duval, S., & Tweedie, R. L. (2000). Trim and fill: A simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 455–463. doi:10.1111/j.0006-341X.2000.00455.x

Ehri, L. C. (2007). Development of sight word reading: Phases and find-

ings. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 135–154). Oxford, United Kingdom: Blackwell.

Fawcett, A. J., & Nicolson, R. I. (1994). Naming speed in children with dyslexia. *Journal of Learning Disabilities, 27,* 641–646. doi:10.1177/002221949402701004

Frost, R. (2005). Orthographic systems and skilled word recognition processes in reading. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 272–295). Oxford, United Kingdom: Blackwell. doi:10.1002/9780470757642.ch15

Georgiou, G. K., Parrila, R., & Kirby, J. (2009). RAN components and reading development from grade 3 to grade 5: What underlies their relationship? *Scientific Studies of Reading, 13,* 508–534. doi:10.1080/10888430903034796

Georgiou, G. K., Parrila, R., Kirby, J. R., & Stephenson, K. (2008). Rapid naming components and their relationship with phonological awareness, orthographic knowledge, speed of processing, and different reading outcomes. *Scientific Studies of Reading, 12,* 325–350. doi:10.1080/10888430802378518

Georgiou, G. K., Parrila, R., & Liao, C. (2008). Rapid naming speed and reading across languages that vary in orthographic consistency. *Reading and Writing, 21,* 885–903. doi:10.1007/s11145-007-9096-4

Georgiou, G. K., Parrila, R., & Papadopoulos, T. C. (2008). Predictors of word decoding and reading fluency across languages varying in orthographic consistency. *Journal of Educational Psychology, 100,* 566–580. doi:10.1037/0022-0663.100.3.566

Hedges, L. V. (1992). Modeling publication selection effects in meta-analysis. *Statistical Sciences, 7,* 246–255. doi:10.1214/ss/1177011364

Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods, 9,* 426–445. doi:10.1037/1082-989X.9.4.426

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3,* 486–504. doi:10.1037/1082-989X.3.4.486

Høien-Tengesdal, I. (2010). Is the simple view of reading too simple? *Scandinavian Journal of Educational Research, 54,* 451–469. doi:10.1080/00313831.2010.508914

Johnston, T. C., & Kirby, J. R. (2006). The contribution of naming speed to the simple view of reading. *Reading and Writing, 19,* 339–361. doi:10.1007/s11145-005-4644-2

Kail, R., & Hall, L. K. (1994). Processing speed, naming speed, and reading. *Developmental Psychology, 30,* 949–954. doi:10.1037/0012-1649.30.6.949

Kail, R., Hall, L. K., & Caskey, B. J. (1999). Processing speed, exposure to print, and naming speed. *Applied Psycholinguistics, 20,* 303–314. doi:10.1017/S0142716499002076

Kairaluoma, L., Torppa, M., Westerholm, J., Ahonen, T., & Aro, M. (2013). The nature of and factors related to reading difficulties among adolescents in a transparent orthography. *Scientific Studies of Reading, 17,* 315–332.

Katzir, T., Kim, Y., Wolf, M., Kennedy, B., Lovett, M., & Morris, R. (2006). The relationship of spelling recognition, RAN, and phonological awareness to reading skills in older poor readers and younger reading-matched controls. *Reading and Writing, 19,* 845–872. doi:10.1007/s11145-006-9013-2

Kirby, J. R., Parrila, R. K., & Pfeiffer, S. L. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology, 95,* 453–464. doi:10.1037/0022-0663.95.3.453

Kirby, J. R., Roth, L., Desrochers, A., & Lai, S. (2008). Longitudinal predictors of word reading development. *Canadian Psychology, 49,* 103–110. doi:10.1037/0708-5591.49.2.103

Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal*

of *Educational Psychology, 100,* 150–161. doi:10.1037/0022-0663.100.1.150

Lervåg, A., Bråten, I., & Hulme, C. (2009). The cognitive and linguistic foundations of early reading development: A Norwegian latent variable longitudinal study. *Developmental Psychology, 45,* 764–781. doi:10.1037/a0014132

Lervåg, A., & Hulme, C. (2009). Rapid automatized naming (RAN) taps a mechanism that places constraints on the development of early reading fluency. *Psychological Science, 20,* 1040–1048. doi:10.1111/j.1467-9280.2009.02405.x

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Newbury Park, CA: Sage.

Mann, V., & Wimmer, H. (2002). Phoneme awareness and pathways into literacy: A comparison of German and American children. *Reading and Writing, 15,* 653–682. doi:10.1023/A:1020984704781

McBride-Chang, C., & Manis, F. R. (1996). Structural invariance in the associations of naming speed, phonological awareness, and verbal reasoning in good and poor readers: A test of the double deficit hypothesis. *Reading and Writing, 8,* 323–339. doi:10.1007/BF00395112

Meyer, M. S., Wood, F. B., Hart, L. A., & Felton, R. H. (1998). Selective predictive value of rapid automatized naming in poor readers. *Journal of Learning Disabilities, 31,* 106–117. doi:10.1177/002221949803100201

Moll, K., Fussenegger, B., Willburger, E., & Landerl, K. (2009). RAN is not a measure of orthographic processing: Evidence from the asymmetric German orthography. *Scientific Studies of Reading, 13,* 1–25. doi:10.1080/10888430802631684

Norton, E. S., & Wolf, M. (2011). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology, 63,* 427–452.

Papadopoulos, T. C., Georgiou, G. K., & Kendeou, P. (2009). Investigating the double-deficit hypothesis in Greek. *Journal of Learning Disabilities, 42,* 528–547. doi:10.1177/0022219409338745

Patel, T. K., Snowling, M. J., & de Jong, P. F. (2004). A cross-linguistic comparison of children learning to read in English and Dutch. *Journal of Educational Psychology, 96,* 785–797. doi:10.1037/0022-0663.96.4.785

Pennington, B. F., Cardoso-Martins, C., Green, P. A., & Lefly, D. L. (2001). Comparing the phonological and double deficit hypotheses for developmental dyslexia. *Reading and Writing, 14,* 707–755. doi:10.1023/A:1012239018038

Protopapas, A., Altani, A., & Georgiou, G. K. (2013). Development of serial processing in reading and rapid naming. *Journal of Experimental Child Psychology, 116,* 914–929. doi:10.1016/j.jecp.2013.08.004

Roman, A. A., Kirby, J. R., Parrila, R. K., Wade-Woolley, L., & Deacon, S. H. (2009). Toward a comprehensive view of the skills involved in word reading in grades 4, 6, and 8. *Journal of Experimental Child Psychology, 102,* 96–113. doi:10.1016/j.jecp.2008.01.004

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. London, England: Sage. doi:10.4135/9781412984997

Savage, R., & Frederickson, N. (2005). Evidence of a highly specific relationship between rapid automatic naming of digits and text-reading speed. *Brain and Language, 93,* 152–159. doi:10.1016/j.bandl.2004.09.005

Savage, R. S., Frederickson, N., Goodwin, R., Patni, U., Smith, N., & Tuersley, L. (2005). Relationships among rapid digit naming, phonological processing, motor automaticity, and speech perception in poor, average and good readers and spellers. *Journal of Learning Disabilities, 38,* 12–28. doi:10.1177/00222194050380010201

Savage, R. S., Pillay, V., & Melidona, S. (2007). Deconstructing rapid automatized naming: Component processes and the prediction of reading difficulties. *Learning and Individual Differences, 17,* 129–146. doi:10.1016/j.lindif.2007.04.001

Scarborough, H. S. (1998). Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness,

verbal memory, rapid naming, and IQ. *Annals of Dyslexia, 48,* 115–136. doi:10.1007/s11881-998-0006-5

Schatschneider, C., Carlson, C. D., Francis, D. J., Foorman, B. R., & Fletcher, J. M. (2002). Relationship of rapid automatized naming and phonological awareness in early reading development: Implications for the double-deficit hypothesis. *Journal of Learning Disabilities, 35,* 245–256. doi:10.1177/002221940203500306

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94,* 143–174. doi:10.1348/000712603321661859

Shen, H. H., & Bear, D. R. (2000). Development of orthographic skills in Chinese children. *Reading and Writing, 13,* 197–236. doi:10.1023/A:1026484207650

Shu, H. (2003). Chinese writing system and learning to read. *International Journal of Psychology, 38,* 274–285. doi:10.1080/00207590344000060

Swanson, H. L., Trainin, G., Necoechea, D. M., & Hammill, D. D. (2003). Rapid naming, phonological awareness, and reading: A meta-analysis of the correlation evidence. *Review of Educational Research, 73,* 407–440. doi:10.3102/00346543073004407

Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine, 22,* 2113–2126. doi:10.1002/sim.1461

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1994). Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities, 27,* 276–286. doi:10.1177/002221949402700503

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997). Contributions of phonological awareness and rapid automatized naming ability to the growth of word-reading skills in second- to fifth-grade children. *Scientific Studies of Reading, 1,* 161–185. doi:10.1207/s1532799xssr0102_4

Vaessen, A., Bertrand, D., Tóth, D., Csépe, V., Faísca, L., Reis, A., & Blomert, L. (2010). Cognitive development of fluent word reading does not qualitatively differ between transparent and opaque orthographies. *Journal of Educational Psychology, 102,* 827–842. doi:10.1037/a0019465

Vaessen, A., & Blomert, L. (2010). Long-term cognitive dynamics of fluent reading development. *Journal of Experimental Child Psychology, 105,* 213–231. doi:10.1016/j.jecp.2009.11.005

Vaessen, A., Gerretsen, P., & Blomert, L. (2009). Naming problems do not reflect a second independent core deficit in dyslexia: Double deficits explored. *Journal of Experimental Child Psychology, 103,* 202–221. doi:10.1016/j.jecp.2008.12.004

van den bos, K. P., Zijlstra, B. J. H., & Spelberg, H. C. I. (2002). Life-span data on continuous-naming speeds of numbers, letters, colors, and pictured objects, and word-reading speed. *Scientific Studies of Reading, 6,* 25–49. doi:10.1207/S1532799XSSR0601_02

van den bos, K. P., Zijlstra, B. J. H., & van den Broeck, W. (2003). Specific relations between alphanumeric-naming speed and reading speeds of monosyllabic and multisyllabic words. *Applied Psycholinguistics, 24,* 407–430. doi:10.1017/S0142716403000213

Vevea, J. L., & Hedges, V. H. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika, 60,* 419–435. doi:10.1007/BF02294384

Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1,* 112–125. doi:10.1002/jrsm.11

Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). Development of young readers' phonological processing abilities. *Journal of Educational Psychology, 85,* 83–103. doi:10.1037/0022-0663.85.1.83

Wimmer, H. (2006). Don't neglect reading fluency! *Developmental Science, 9,* 447–448. doi:10.1111/j.1467-7687.2006.00527.x

Wimmer, H., Mayringer, H., & Landerl, K. (2000). The double deficit hypothesis and difficulties in learning to read a regular orthography. *Journal of Educational Psychology, 92,* 668–680. doi:10.1037/0022-0663.92.4.668

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology, 91,* 415–438. doi:10.1037/0022-0663.91.3.415

Wolf, M., Bowers, P. G., & Biddle, K. (2000). Naming-speed processes, timing, and reading: A conceptual review. *Journal of Learning Disabilities, 33,* 387–407. doi:10.1177/002221940003300409

Wolf, M., O'Rourke, A. G., Gidney, C., Lovett, M., Cirino, P., & Morris, R. (2002). The second deficit: An investigation of the independence of phonological and naming speed deficits in developmental dyslexia. *Reading and Writing, 15,* 43–72. doi:10.1023/A:1013816320290

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., . . . Blomer, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science, 21,* 551–559. doi:10.1177/0956797610363406

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131,* 3–29. doi:10.1037/0033-2909.131.1.3

Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science, 12,* 379–384. doi:10.1111/1467-9280.00370