Validating Phenomenological Aspects of the Mental Imagery Experience through Meta-Analysis: Beyond Global Assessment of Imagery Ability

Matthew S. Runge

Carleton University 2012/2013
Abstract

The phenomenological experience of mental imagery vividness is of increasing interest within the field of mental health science, yet commentators refute its empirical validity as an independent scientific construct. Two assessments of mental imagery vividness were compared through meta-analysis, the Vividness of Visual Imagery Questionnaire (VVIQ), and trial-by-trial ratings of vividness. Each vividness assessment was further divided into behavioural/cognitive or neuroscientific measures. A corpus of 965 peer-reviewed journal articles were retrieved from four major databases and relevant statistical outcomes from each paper were recorded. Effect size estimates were computed for 3579 statistical outcomes, which were categorized as into one of four comparison groups (Vividness and Behavioural/Cognitive, VVIQ and Behavioural/Cognitive, Vividness and Neuroscientific, and VVIQ and Neuroscientific). It was found that the average effect size magnitude for trial-by-trial vividness exceeded that of the VVIQ for behavioural/cognitive, but not neuroscientific measures. However, the average effect sizes magnitude for neuroscientific measures was generally greater than behavioural/cognitive ones. Additionally, the average effect sizes magnitude for trial-by-trial vividness ratings was generally greater than the VVIQ. It is suggested that trial-by-trial ratings, in conjunction with neuroscientific measurement, may provide a more precise and reliable measure of mental imagery vividness. Despite face validity, unique observations correlating trial-by-trial vividness ratings with the VVIQ were weak to moderate on average. Theoretical considerations on the empirical validity of the construct of vividness are discussed.

Keywords: VVIQ, vividness, neuroimaging, meta-analysis
Acknowledgements

The author would like to take this opportunity to personally acknowledge the many and mighty contributors whom facilitated the current research—may the force be with them always. First and foremost, the persistent and untiring guidance of my supervisor, colleague and friend Dr. Amedeo D’Angiulli has been indispensible. Thank you for your courageous insights, and persistent teaching. For the unbelievable precedent established in the early stages of this project—a hat tip to Aldrich Chan and Val Bakhilau. For their steadfastness and professional example—a special thanks to my committee members Alfonso Abizaid and Matthew Holahan. Additional thanks to Shawn Hayley and Jim Davies for their participation in this project as well. To all members of the scientific community, I have said it before, and I’ll say it again:

“Did you know Voltaire was the first to suggest that the universe was created by a gigantic explosion..., and that Goethe was the first to suggest that spiral nebulae were swirling masses of stars?... We now call them galaxies. It is kind of funny how often new concepts of science find their first tenets and forms of expression in the arts. So..., did Beethoven do physics on the side”?

The Man from Earth (Wilkinson, 2007)
# Table of Contents

Abstract..................................................................................................................................................2

Acknowledgements..................................................................................................................................3

Table of Contents..................................................................................................................................4

List of Tables.........................................................................................................................................6

List of Figures.......................................................................................................................................7

List of Equations...................................................................................................................................8

List of Appendices...............................................................................................................................9

Validating Phenomenological Aspects of the Mental Imagery Experience through Meta-
Analysis: Beyond Global Assessments of Imagery Ability....................................................................10

  *Historical Background....................................................................................................................10*

  *Mental Imagery as an Adjunct in the Treatment of Neuropathology..............................................14*

  *The Present Study..........................................................................................................................20*

Method..................................................................................................................................................24

  *Sampling Methodology..................................................................................................................24*

  *Data Trimming and Statistical Analysis.........................................................................................27*

Results..................................................................................................................................................29

  *Descriptive Statistics and Assumptions.........................................................................................29*
List of Tables

Table 1: Descriptive statistics for comparison groups (untransformed absolute Z). .................. 30

Table 2: Descriptive statistics for comparison groups (transformed absolute Z). .................... 31

Table 3: Descriptive statistics for comparison groups (transformed absolute Z, no outliers). ..... 34
List of Figures

Figure 1: Prototypical paradigm for measuring vividness on a trial-by-trial basis.........................12

Figure 2: Mean Transformed Z for Rating collapsed across Measure...........................................33

Figure 3: Mean Transformed Z for Measure collapsed across Rating...........................................34

Figure 4: Mean Transformed Z for Rating collapsed across Measure...........................................37

Figure 5: Mean Transformed Z for Measure collapsed across Rating...........................................38
List of Equations

Equation 1: Fisher’s transformation for effect size $r$ ................................................................. 28

Equation 2: Calculation of Fail-safe N for Cohen’s $d$ ................................................................. 45

Equation 3: Calculation of Cohen’s $d$ from effect size $r$ for Equation 2 .................................. 45
List of Appendices

Appendix A: The VVIQ.................................................................................................................................71

Appendix B: Formulae for the Conversion of Test Statistics to Standard Effect Size $r$.................72

Appendix C: Complete Meta-Analysis Reference List..................................................................................73
Validating Phenomenological Aspects of the Mental Imagery Experience through Meta-Analysis: Beyond Global Assessments of Imagery Ability

**Historical Background**

Mental images or mental imagery refers to the subjective experience of a percept in the absence of a relevant physical stimulus (Hebb, 1968). In spite of the fact that there have been major debates about the underlying nature and the specific format of imagery (e.g., Pylyshyn, 2002; Kosslyn, Thompson & Ganis, 2006), in this context these debates are irrelevant since no one denies the subjective phenomenological experience associated with imagery, which is the aspect on which the present work focuses. According to Galton (1880), the detail and clarity of the experience associated with imagery is variable, and operationally defined as the vividness of the event. Historically, the breadth of mental imagery includes, but is not limited to the visual modality. Betts (1909) devised the Questionnaire upon Mental Imagery (QMI), and addressed the phenomenology of vividness along seven sensory dimensions (visual, auditory, olfactory, gustatory, cutaneous, kinaesthetic, and organic), which were measured through a 150 item inventory. Responses were to be recorded along a seven-point scale from “perfectly clear and as vivid as the actual experience” to “no image at all.” The summated score offered an index of imagining ability. Although the questionnaire was considered a sufficient measure of imagery ability, the extensiveness of the QMI rendered it administratively cumbersome in clinical and experimental application (Sheehan, 1967). Subsequently, the QMI underwent significant excision. Reducing the number of questions to 35 decreased the laborious amount of time required to administer the questionnaire, and increased the feasibility of imagery diagnostics in research settings. However, as the need for a succinct and timely questionnaire developed, so too did the need for a modality-specific one.
The need for a modality specific assessment of visual imagery ability was the impetus behind the advent of the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973; see Appendix A). The VVIQ provides a global assessment of vividness, and is typically used to segregate participants according to individual differences in visual imagery ability. It consists of 16 items, which are to be rated on a five-point scale, where (1) “perfectly clear and as vivid a normal vision,” (2) “clear and reasonably vivid,” (3) “moderately clear and vivid,” (4) “vague and dim,” (5) “no image at all, you only ‘know’ that you are thinking of the object.” The versatility of the VVIQ is such that it can be administered before, during, or after experimental manipulations, with sufficient retest and internal reliability (Campos, Amor, & Gonzalez, 2002). In experimental contexts, vividness is almost exclusively measured using the VVIQ (Pearson, 1995). The VVIQ was later revised (VVIQ2) to include twice as many questions (Marks, 1995). However, in accordance with Sheehan’s reasoning, the feasibility of measuring imagery vividness in experimental situations may more easily be afforded through a succinct and timely assessment. An alternative approach to assessing imagery ability is to employ a rating scale similar to the one used in the VVIQ throughout an experimental procedure, such that participants rate the vividness of their image in each trial, immediately following image formation.

Imagery vividness may be rated on a trial-by-trial basis through a single response, which corresponds to the subjective experience at any particular moment in time. In addition to methodological advantages, Hertzog and Dunlosky (2006) argue that trial-by-trial vividness ratings are perhaps the most effective means by which imagery vividness can be studied. For example, D’Angiulli and Reeves (2003-2004) had participants fixate on the uppermost side of a square presented in the center of a computer screen. Participants initiated each trial by clicking a mouse, at which point an alerting beep was sounded, followed 250msec later by a single noun
word (e.g. zebra) appearing at above the square. Participants were asked to read the word silently, and form a mental image of the word. Upon completely resolving the mental image, participants were informed that they were to respond with a mouse-button press, at which point participants were prompted with a seven-point vividness rating scale presented from left to right across the screen. From left to right, the options were, (1) “no image,” (2) “very vague/dim,” (3) “vague/dim,” (4) “not vivid,” (5) “moderately vivid,” (6) “very vivid,” (7) “perfectly vivid”). Participants indicated the perceived vividness of their mental image by pressing the mouse-button along the corresponding vividness rating. It was emphasized to participants in advance that ratings of vividness referred to the degree of clarity and detail inherent in the mental imagery experienced. These options were (1) “no image present at all, you only “know” that you are thinking of the object,” (2) “so vague and dim as to be hardly discernable,” (3) “vague and dim,” (4) “not clear or vivid, but recognizable,” (5) “moderately clear and vivid,” (6) “very clear and comparable in vividness to the actual experience as real seeing,” (7) “perfectly clear and as vivid as the actual experience as real seeing”. This prototypical paradigm for measuring vividness self-reports on a trial-by-trial basis is presented in Figure 1.

![Figure 1](image)

*Figure 1.* Prototypical paradigm for measuring vividness on a trial-by-trial basis. Options for vividness ratings are (1) “no image,” (2) “very vague/dim,” (3) “vague/dim,” (4) “not vivid,” (5)
“moderately vivid,” (6) “very vivid,” (7) “perfectly vivid”). Figure adapted from D’Angiulli (2001).

Global assessments offered through surveys of group differences in mental imagery ability may trivialize the cognitive and neural processes affiliated with the phenomenological experience of vividness occurring within each trial. Such a conceptualization can be likened to “trait” and “state” vividness, whereby global assessments resolve a general proficiency concerning imagery ability, and trial-by-trial ratings resolve situational differences in the mental imagery experience, respectively. Indeed, experimental procedures employing trial-by-trial vividness ratings may procure an entirely different pattern of results and interpretations than those employing the VVIQ (D’Angiulli, 2002; D’Angiulli & Reeves, 2002, 2007; Alter & Balcetis, 2010; Rabin, Gilboa, Stuss, Mar, & Rosenbaum, 2010; Pearson, Rademaker, & Tong, 2011). If ratings of vividness on a trial-by-trial basis are observed in structured experimental settings, and participants are given clear instructions as to the nature of the task, vividness ratings may resolve critically informative aspects concerning the subjective experience of the imagery processes (Baddeley & Andrade, 2000; Giusberti, Cornoldi, de Beni, & Massironi, 1992; Hale & Simpson, 1971). Trial-by-trial vividness ratings, then, can be considered a type of retrospective verbal report (Ericsson & Simon, 1993), which can be related to task behaviour and neural processes. Moreover, the validation of vividness as a construct is largely impossible through the use of a single test score, such as the one offered by the VVIQ or global assessments of imagery ability more generally.

The legitimacy of vividness as an objective scientific construct has been subjected to significant empirical scrutiny (McKelvie, 1995; Pylyshyn, 2003). Although novel neurocognitive imaging techniques have facilitated researcher understanding of the underlying neurophysiology
of mental imagery (Cichy, Heinzle, & Haynes, 2012), the phenomenological correlates have undergone extensive refute; albeit, opponents have not been convincingly successful as of late (Millar, 1994; Pylyshyn, 2003). The construction of vividness is contingent on convergent validation with more objective criteria, as evidenced by positive correlations with performance on memory tasks (Baddeley & Andrade, 2000), arousal level (Barrowcliff, Gray, Freeman, & Macculloch, 2004; Bywaters, Andrade, & Turpin, 2004), and sleep stages (Conduit, Crewther, & Coleman, 2004). Although vividness appears to be unrelated to certain dynamic spatial tasks like mental rotation (Dean & Morris, 2003), it does not necessarily negate the validity of the measure. Rather, the apparent lack of a relationship may imply that visual imagery and dynamic imagery are unrelated processes, which may serve to foster the tenability of the vividness construct by demonstrating discriminant validity. Moreover, since vividness ratings have been shown to correlate with increased visual cortex activity relative to the rest of the brain (Gonsalves & Paller, 2002; Cui, Jetter, Yang, Montague & Eagleman, 2007), neurocognitive imaging research may provide an objective index via neuroscientific measures. Neuroimaging may resolve aspects of phenomenological experience which is concealed to behavioural and cognitive testing, leading to application in mental health research and practice. Clinical interventions implementing mental imagery strategies often offer a distinct prognostic advantage for a variety of psychological disorders, one which is equally effective, and typically less intrusive than alternative treatment options (Reardon & Tosi, 1977; Hunt et al., 2006). Furthermore, in conjunction with neuroimaging techniques, mental imagery strategies also provide health care professionals the opportunity to improve the quality of patient care and comfort (Owens et al., 2006; Wilhelm, Jordon, & Birbaumer, 2006).

Mental Imagery as an Adjunct in the Treatment of Neuropathology
Through the characterization of the vividness construct as a self-report measure, it is the objective of the present research to establish not only a methodological precedent, but delineate the relevance of mental imagery in the mental health field. Because subjective self-reports lack physical dimensions, active streams of scientific research have deprived the vividness construct of sincere systematic scrutiny, reduced it to one of epiphenomenal significance, and perhaps underestimated its empirical validity. Albeit, the current opinion in mental health research is one which seeks to revive the active pursuit of subjective vividness, in an effort to both understand the fundamental properties of the brain, and optimize clinical treatment strategies. The explicit measurement of vividness as a research prerogative may provide information critical to the theoretical understanding of mental illness, and the way in which it may influence related psychological dysfunction. For example, one’s ability to imagine may mediate physiological responses (Cook, Melamed, Cuthbert, McNeil, & Lang, 1988; Tiffany & Drobes, 1990; Tiffany & Hakenewerth, 1991; McNeil, Vranan, Melamed, Cuthbert, & Lang, 1993; Johnsen & Lutgendorf, 2001; Van Diest et al., 2001; Van Diest et al., 2005).

The experience of vividness can be significantly altered through neurochemical modulation, which may facilitate psychotherapy. For example, the administration of nitrous oxide improves imagery vividness and imaginative suggestibility (Whalley & Brooks, 2009), which may compound the effectiveness of hypnosis, analgesia, and conscious sedation. Carhart-Harris and colleagues (2012) demonstrated that ratings of memorial vividness and mental imagery were greater for subjects receiving psilocybin than controls, which correlated quite strongly with subjective well-being, and may enhance autobiographical recollection in clinical settings. Conversely, the neurochemical modulation of vividness may be disruptive to physiological processes. Nicotine patches were observed to increase the quantity of dream
reports containing visual imagery, and the vividness of said dreams during REM sleep in a sample of smokers, which was positively related to increased sleep fragmentation and less REM sleep (Page, Coleman, & Conduit, 2006). Although beta-adrenergic blocking agents sufficiently manage cardiovascular disorders, they have been documented to cause vivid imagery, nightmares and hallucinations (Rosenman & Maser, 1999).

The impetus behind the resurgent interest in mental imagery in mental health science is the seemingly concomitant relationship between mental imagery and neuropathology characterized by emotional dysfunction (Shin, Rauch, & Pitman, 2006; Soderlund, Moscovitch, Kumar, Mandic, & Levine, 2012). Post traumatic stress disorder (PTSD) is perhaps the most exemplary clinical criteria, wherein vivid imagery and emotional dysregulation partially satisfies a differential diagnosis (4th ed., text rev.; DSM–IV–TR; American Psychiatric Association, 2000). Sometimes, and perhaps inappropriately referred to as “flashbulb” memories, PTSD is characterized by intrusive, vivid, and involuntary retrieval of traumatic events, which are typically perceived through the visual modality, and accompanied by negative affect such as fear, helplessness, guilt, anger, shame and horror (Hackmann, 2011). Treatment strategies include elaboration of the traumatic stimulus into a coherent autobiography, cognitive reappraisal, and the desistance of suppression, avoidance and rumination cognitions. However, an alternative clinical intervention mediating the relationship between mental imagery and emotional processing in PTSD is eye movement desensitization and reprocessing (EMDR; Shapiro, 1989; Andrade, Kavanagh, & Baddeley, 1997), the implementation of which has been successful in the treatment of depression (Brewin, 1998; Ironson, Freund, Strauss, & Williams, 2002), and appetitive disorders (McClelland, Kems, & Tiggemann, 2006), and may be more generally applicable to various psychopathologies wherein vivid and intrusive mental imagery
are symptomatic (Brewin, 2010). Although in some studies eye-movements have been shown to have no effect, or increase imagery vividness in certain situations (Neisser, 1967 (p. 153); Hale & Simpson, 1970), and eye-movements during imagery may represent the content of the visualized scene (Brandt & Stark, 1997), EMDR has received extensive clinical support (van den Hout & Engelhard, 2012).

The EMDR intervention requires patients to repetitiously imagine a traumatic experience, while concurrently engaging in eye movements. The systematic association between the automaticity of the intrusive image, and the sustained cognitive effort of visual tracking eventually fosters a reduction in emotional saliency and image vividness; thereby, relieving patient distress. Gunter and Bodner (2008) posit cognitive resource theory may most aptly explain the mechanism by which EMDR operates. The cognitive effort affiliated with maintaining eye movements taxes the visuospatial subsystem of working memory, and disproportionately consumes attentional resources otherwise utilized in vivid imagining and emotional processing. Extending this finding, Smeets, Dijs, Pervan, Engelhard, and Van den Hout (2012) provide strong evidence that decreases in imagery vividness precede those of emotionality. These researchers randomly assigned participants to eye stationary or eye movement conditions, and had participants recall unpleasant memories over a period of 96 seconds. Emotionality and vividness ratings were probed on a trial-by-trial basis at regular intervals throughout the procedure. It was found that both sets of ratings steadily decreased in the eye movement condition relative to the eye stationary condition, but decreases in vividness ratings always accompanied or preceded decreases in emotionality. With respect to EMDR therapy, these data may imply changes in phenomenological aspects of the imagery experience.
are necessary and sufficient causal antecedents to changes in emotionality, and may occupy a more fundamental stage in processing.

Although neurobiological theories depicting the mechanism whereby EMDR operates vary widely in their proposed neurophysiology (van den Hout & Engelhard, 2012), the anterior cingulate cortex (ACC) resides at the crux of these models (Bergmann, 2010). For example, reciprocal suppression/inhibition models of EMDR (Corrigan, 2002; Kaye, 2007) specifically implicate the ACC in the regulation of cognitive and affective processes through two separate routes of processing, which are consistent with morphologically defined subdivisions (Devinsky, Morrell, & Vogt, 1995; Bush, Luu, & Posner, 2000). The affective subdivision of the ACC (aACC) extends ventrally throughout the limbic system and adjacent structures, such as the amygdala, nucleus accumbens, hippocampus, hypothalamus, and autonomic brainstem motor nuclei. The aACC contributes to the monitoring of performance, internal physiological states, and the presence of rewards and punishments (Davidson et al., 2002; Whalen et al., 1998). The cognitive subdivision (cACC) principally extends along the dorsal dimension of the ACC towards posterior structures, such as the posterior cingulate cortex (PCC), and parietal lobe. The cACC contributes to the modulation of attentional resources, executive functions, and effortful control (Davidson et al., 2002; Bush et al., 1998). According to suppression/inhibition models (Corrigan, 2002; Bergmann, 2010) eye-movements bilaterally stimulate thalamo-cingulate tracts, and selectively deactivate the affective subdivision, which then activates the cognitive subdivision through reciprocal inhibition.

Within the cognitive subdivision, direct connections from the PCC unto the precuneus likely corresponds to the phenomenological experience of conscious information processing (Vogt & Laureys, 2005), and mental imagery more specifically (Fletcher et al., 1995; for a
review see Cavanna & Trimble, 2006). Petit and colleagues (1993) measured regional cerebral blood flow (rCBF) in right handed individuals while self-paced horizontal, voluntary saccadic eye movements were made in complete darkness. Increased subcortical blood flow was observed in the putamen, globus pallidus, and thalamus, and increased cortical blood flow was observed in the precentral gyrus, and superior median frontal gyrus, during regular eye movements. Significant activation was also documented in the superior ACC and PCC regions, the insula, and the cerebellum, as were weak activations in the fusiform and lingual gyri. Because thalamic nuclei participate in the control and execution of saccadic eye movements (Schlag-Rey & Schlag, 1989), the interaction between the thalamus and areas of the cACC and PCC may coincide with the site of EMDR action. Because changes in vividness precede those of emotionality (Smeets et al., 2012), one would predict normal cACC functioning to be decremented prior to aACC functioning, which directly belies the framework proposed by proponents of the suppression/inhibition model.

In comparison to the aACC, the cACC may be the pathway most modulated by thalamic projections, and the oculomotor information they transmit. Vogt, Finch, and Olsen (1992) describe the location wherein oculomotor signals converge within the PCC as the visuospatial area. Neurons within the visuospatial area are variably encoded for the size and direction of an eye movement, and also the position of the orbit along the vertical dimension. The configuration of eye movements is perhaps most disrupting to neuronal transmission about the cACC route of processing. Neurons specialized for saccadic movements may interfere with normal PCC functioning; thereby reducing the phenomenological experience of vividness. Although eye movements may tax the visuospatial subsystem of working memory (Gunter & Bodner, 2008), it
is perhaps more accurate to presume eye movements interfere with the transmission of information within the visuospatial subsystem, a process which later effects emotionality.

Despite over 20 years of extensive investigation (Bergmann, 2010), the neurobiological mechanism underlying EMDR remains elusive (van den Hout & Engelhard, 2012). However, contemporary neurobiological theorizations may trivialize the contribution of phenomenological vividness within the context of the therapy (Corrigan, 2002; Harper, Rasolkhani-Kalhorn, Drozd, 2009). The present neurobiological conceptualization of EMDR suggests that the thalamus may modulate the phenomenological experience of vividness at the level of mental imagery formation, reductions of which occur prior to those of emotionality. As such, the measurement of vividness may provide critical information concerning the therapeutic basis and effectiveness of EMDR therapy, the content of which global assessments of imagery ability may insufficiently resolve.

**The Present Study**

In summary, the preceding portions of this narrative systematically delineated the study of mental imagery vividness from the historical to the contemporary, and described the ways in which vividness can, and has been studied. This narrative has sought to describe the relationship between vividness and neuropathology characterized by differential mental imagery functioning, such as PTSD. Upon describing the relevance of mental imagery in the realm of clinical neuroscience, the abovementioned paragraphs speculate on the neurophysiologic relationship between phenomenological vividness and affective processes. The purpose of the current paper is designed to demonstrate the relevance of phenomenological vividness in the realm of clinical
neuroscience, refine the way in which it may be studied, and provide a point of reference for future research.

The proposed research is one of synthesis, analysis, and evaluation, and represents the preliminary step to enable a test of the approach put forward to the previous paragraphs in a much larger research program. Although mental imagery vividness and emotional intensity may be concomitantly related and relevant to clinical neuroscience more generally, the preliminary phase of the program is a descriptive analysis of the available literature pertaining to vividness as an empirical construct. In this manner, the present study seeks to validate the study of vividness as an independent construct, which is to be compared against all available behavioural, cognitive, and neuroscientific measures throughout a robust, representative sample of literature. Using meta-analytic procedures, in conjunction with statistical modelling, the preliminary phase of the current research has been designed to empirically test the contention that self-reports of mental imagery vividness dispossess scientific integrity. Because the experimental paradigms almost exclusively measure group differences in imaging ability (Pearson, 1995), the VVIQ is to provide the standardized metric against which self-reports will be compared and validated.

Despite the significant empirical refute surrounding the study of mental imagery more generally (McKelvie, 1995; Pylyshyn, 2003), trial-by-trial self-reports may resolve critical aspects underlying cognitive and neural processes, the likes of which global assessments are insufficiently sensitive enough to measure (Hertzog & Dunlosky, 2006). Moreover, the validation of vividness as a construct is largely impossible through the use of a single test score. Although the VVIQ and trial-by-trial ratings may converge to validate the construct of vividness, trial-by-trial ratings may demonstrate greater content validity, and more sufficiently represent every element of the mental imagery process. As mentioned previously, the concurrent
VALIDATING PHENOMENOLOGICAL ASPECTS

utilization of both measures typically results in disagreement (D’Angiulli, 2002; D’Angiulli & Reeves, 2002, 2007; Alter & Balcetis, 2010; Rabin, Gilboa, Stuss, Mar & Rosenbaum, 2010; Pearson, Rademaker & Tong, 2011). Therefore, the preliminary phases of the current research are synthetic, designed to survey the available literature, and describe the relationship between the VVIQ and trial-by-trial ratings against more objective criteria. Meta-analysis is a collection of procedures for systematically reviewing, integrating, and comparing research of a common interest (see Cooper and Hedges, 1994; p. 6). It is in this manner that the relative effect size for the VVIQ and trial-by-trial vividness ratings can be exhaustively compared, over the corpus of literature utilizing these measures of interest, which may be obtained through a variety of available databases. The results are to be summarized, compared, and contrasted as a series of planned comparisons, on the basis of the unstandardized effect size metric ($r$; Cohen, 1988). Because trial-by-trial vividness ratings resolve critical aspects of, and measure a greater subset of the variability in the mental imagery experience, they may systematically demonstrate greater content validity than the VVIQ within behavioural and cognitive experimental paradigms (Smeets et al. 2012; Baddeley & Andrade, 2000; Giusberti, Cornoldi, de Beni, & Massironi, 1992; Hale & Simpson, 1971). Accordingly, one would expect larger effect sizes for vividness ratings for research of this type. Richardson (1994) provides a supporting view. If trial-by-trial vividness ratings sufficiently resolve phenomenological aspects of mental imagery to a greater degree than the VVIQ, then the reported effect sizes pertaining to trial-by-trial vividness ratings should be greater in magnitude, when averaged over the population of literature pertaining to behavioural and cognitive measures. Although self-reports of vividness can be accrued for virtually any sensory modality, visual mental imagery is perhaps most relevant to clinical
neuroscience, and represents a majority of the available research. Therefore, the present argument is limited to that of visual mental imagery.

Included in the current research is the secondary phase of the program. The secondary phase of the program is one which is designed to model the data collected in the preliminary phase of the program, and derive inferential predictions based on the type of measure against which vividness is compared. The viability of an empirical based approach to mental imagery is one which is theoretically supported through neurophysiologic modelling (Cichy, Heinzle & Haynes, 2012), and neuroimaging experiments typically exhibit a great amount of control and precision (Kosslyn, 1994). The neural basis to mental imagery is a subset of brain system involved in perception (Amedi, 2005; Ganis, Thompson, & Kosslyn, 2004), and self-reports of vividness on a trial-by-trial basis do correlate with transient brain activity under certain circumstances (Gonsalves & Paller, 2002; Cui, Jetter, Yang, Montague, & Eagleman, 2007). Given neuroscientific measures may sufficiently resolve critical aspects of the mental imagery experience more closely than behavioural and or cognitive ones, this finding may foster construct validity for vividness, and demonstrate the independence of vividness as an empirical measure. If neuroscientific measures are capable of resolving phenomenological vividness more clearly than behavioural and or cognitive ones, then the reported magnitude of effect sizes which are neuroscientific should be greater than those which are behavioural and or cognitive. Specifically, trial-by trial measures of vividness should also be greater than the VVIQ for effect sizes which are neuroscientific.

Upon the completion of the current research, the tertiary phase of the proposed research program is designed to integrate and apply the knowledge accumulated throughout the preliminary and secondary phases into mainstream research, and optimize the way in which
mental imagery vividness is studied. As per the preliminary and secondary phases, the validation of the vividness construct may then be systematically related to other measures relevant to clinical neuroscience, such as those outlined in the preceding paragraphs of this text. As per the secondary phase, neurophysiologic testing may sufficiently resolve critical aspects of the cognitive and neural processes involved in the mental imagery process. Taken together, the first two phases of this program may serve as a point of reference for future research, which is to serve as an origin for imminent experiments involving human participants, and possibly sensitive populations. Such content, construct and convergent validity would sufficiently establish phenomenological aspects of the mental imagery experience as legitimately scientific; thereby, empirically validating the construct. If correct, the present meta-analysis should foster novel research directions, and theoretical advances predicated on self-reported phenomenological experience.

Method

Sampling Methodology

A corpus of peer-reviewed journal articles representing a robust subset of the relevant literature available through the databases Web of Science, Scopus, Embase, and PsycINFO were retrieved through the Carleton University library on October 25\textsuperscript{th}, 2012. Email alerts were enabled such that relevant publications fitting the search criteria, yet appearing in press after this date could be introduced into the analysis on an ex post facto basis. A priori criteria restricted the search results to those of the English language, to those published after 1950, and those using human subjects. Theses, dissertations, and unpublished works were deliberately avoided. Preliminary search parameters including the terms “vivid* and image*” rendered a surplus of papers unrelated to the
research question (note: the asterisk function permits any adjunct suffix; i.e. “vividly”, “vividness”). In an effort to reduce the proportion of false-positive results, the aforementioned search terms were amended with the syntax “NOT haptic NOT ultras* NOT atrial NOT echo* NOT flavor* NOT rats NOT motor NOT auditory NOT surface* NOT arter* NOT olfac*,” as these outcomes were observed to be relatively frequent, however unrelated. Upon adjusting the search parameters to minimize false-positive contamination, 1289 journal articles (Web of Knowledge, 36; Scopus, 34; Embase, 188; PsycINFO, 1032) were exported to RefWorks, wherein 324 duplicates were observed and deleted from the database. Given initial searches rendered a quantity of papers exceeding 10,000, such a sample size was decidedly both sufficiently large, yet sizably manageable.

The remaining 965 papers were retrieved from the Carleton and Ottawa University libraries, and systematically reviewed by a Graduate student cognizant of the research question, and experienced with research synthesis methodology (for complete reference list see Appendix C). Upon investigating each paper for statistical outcomes directly relating to the VVIQ and or trial-by-trial ratings of vividness occurring through the visual modality, 573 papers were dismissed. Of these, 80 were identified as theoretical papers and 493 utilized measures outside of, or failed to include data pertaining to the scope of the present investigation. For example, Kendall, Hrycaiko, Martin, and Kendall (1990) reported using the VVIQ, yet provided no statistical outcomes pertaining to it. Measures outside of scope of the present investigation included, but were not limited to alternative questionnaires concerning mental imagery, such as the Questionnaire upon Mental Imagery (QMI; Sack, van de Ven, Etschenberg, Schatz, & Linden, 2005), the Movement Imagery Questionnaire (MIQ; Wilson, Smith, Burden, and Holmes, 2010), Vividness of Imagery Questionnaire (VIQ; Picardô, Lebaz, Jouffrais, &
Monnier, 2010), the Verbalizer-Visualizer Questionnaire (VVQ; McGrath, O'Malley, Dura, & Beaulieu, 1989), the Object-Spatial Imagery Questionnaire (OSIQ; Blajenkova, Kozhevnikov, & Motes, 2006), the Vividness of Visual Imagery Test (VVIT; Campos, 2009), the Imagery Vividness Questionnaire (IVQ; Campos, Perez-Fabello, & Gomez-Juncal, 2006), and the Gordon Test of Visual Imagery Control (GTVIC; Golomer, GravenHorst, & Toussaint, 2009). Variations to the VVIQ, such as the Vividness of Movement Imagery Questionnaire (VMIQ; Taktek, Zinsser, & St-John, 2008), the Vividness of Olfactory Imagery Questionnaire (VOIQ; Stevenson, Case, & Mahmut, 2007), the Kinaesthetic and Visual Imagery Questionnaire (KVIQ; Dettmers, Benz, Liepert, & Rockstroh, 2012), the Vividness of Visual Imagery for Parents Questionnaire (VVIPQ; McKelvie, 1998), the Auditory Imagery Questionnaire (AIQ; Hishitani, 2009), the Vividness of Haptic Mental Imagery Questionnaire (VHMIQ; Campos, López, Pérez, 1998) and Ashen Ashen’s Adapted Vividness of Visual Imagery Questionnaire (AA-VVIQ; Ashen, 1993) were likewise considered outside of the scope of the present investigation.

Trial-by-trial ratings of vividness operating through sensory modalities other than visual, such as auditory (Khatena, 1976), and kinaesthetic (MacIver, Lloyd, Kelly, Roberts, & Nurmikko, 2008), or operating through a combination of sensory modalities simultaneously (Kemps & Tiggemann, 2007) were considered outside of the scope of the present investigation. As were vividness ratings pertaining to abstract referents such as figural, mimetic and symbolic imagery (Cartwright, 1980), and vividness ratings pertaining to processes other than mental imagery, such as perceptual (Zanker, Doyle, & Walker, 2003) or memorial (Wright, Gaskell, O'Muircheartaigh, 1998) vividness. Vividness measurements other than those subjectively registered by participants were also considered outside of the scope of the present investigation,
such as the methodology employed by Hunt and colleagues (2006), wherein five independent judges rated the vividness and horror of descriptions made by snake phobic individuals.

No data relating to factor loadings (Childers, Houston, & Heckler, 1985), alpha reliabilities (Howie, Markham, & Kleitman, 2006), split-half, test-retest, nor parallel form reliabilities (McKelvie & Gingras, 1974), were considered. Data from the remaining 392 papers, which included at least one measure of interest, were analysed, wherein any and all statistical outcomes relating to the measures of interest were extracted, and entered into a database in the form of test statistic (i.e. F-values, t-values, $\chi^2$-values, Z-values, etc.), degrees of freedom, sample size, and probability signifiers (Friedman, 1982; see Appendix B for conversion formulae). These data were utilized in the computation of the unstandardized effect size ($r$), before being subject to further processing and statistical analysis, the utilization of which offers distinct advantages over $d$ in meta-analysis (Rosenthal & DiMatteo, 2001). Irrespective of relevance to the present investigation, statistical outcomes which provided only probability signifiers were decidedly ignored a priori, as they do not provide sufficient information such that a standard effect size can be computed with any degree of accuracy (i.e. $p < 0.01$, $p > 0.10$, etc.).

Upon entry, each statistical outcome was assigned to one of five groups (“Vividness and Behavioural/Cognitive”, “VVIQ and Behavioural/Cognitive”, “Vividness and Neuroscientific”, “VVIQ and Neuroscientific”, and “Vividness and VVIQ”).

Data Trimming and Statistical Analysis

A total of 3687 observations were made as per the aforementioned criteria, of which 108 were deleted under the following proviso. Of these, 35 entries did not provide a test statistic in conjunction with the corresponding degrees of freedom, nor could degrees of freedom be
confidently estimated, such that a unstandardized effect size could be calculated. For example, Keng and Lin (2006) provided a series of statistical outcomes relevant to trial-by-trial ratings of vividness, yet did not provide the corresponding degrees of freedom for the F-values they reported. If, however, a statistical test was reported without the corresponding degrees of freedom, but degrees of freedom could be estimated without conjecture (for example, Chen & Williams, 2012), the missing value was substituted for N-2 (where N is equivalent to the number of participants involved in the comparison). Estimates were estimated for 88 observations. Test statistics reported as <1, of which there were 85, were replaced with 0.50, and χ² values which rendered an r value greater than 1, of which there were 3, were replaced with 0.99, such that Zᵣ value could be computed (see Equation 1). Beta coefficients were not considered in the present analysis, of which there were 73 observations. Although Beta coefficients residing within the interval ±0.50 can be used interchangeably with values of r (Peterson & Brown, 2005), a total of 7 observations exceeded this interval, and it was thought that the inclusion of a subset of the Beta coefficients would bias the results. In addition to inflating the standard effect size for values residing outside the ±0.50 interval, if Beta coefficients were used to impute standard effect sizes, then a Zᵣ value could not be computed for values greater than, or less than ±1, respectively.

\[ Zᵣ = \frac{1}{2} \ln \left( \frac{1 + r}{1 - r} \right) = artanh(r) \]

Equation 1: Fisher’s transformation for effect size r. The transformation is identical to the inverse hyperbolic tangent of r. For r values greater than 1, and less than -1, no computation can be performed, as this is an undefined operation.

A total of 3579 statistical outcomes were observed from the 392 papers examined (Vividness and Behavioural/Cognitive, 1760; VVIQ and Behavioural/Cognitive, 1629; Vividness...
and Neuroscientific, 80; VVIQ and Neuroscientific, 63; and Vividness and VVIQ, 47). For each of these observations, an individual $Z_r$ transformation was computed. Of unique interest in the present analysis were observations directly comparing trial-by-trial ratings of vividness with the VVIQ (Vividness and VVIQ). Therefore, these data were analysed in isolation, were not included in statistical analyses in conjunction with the four comparison groups, and were considered separately. Data were analysed using nonparametric and parametric analyses. All data were organized using Microsoft EXCEL 2007, and analyzed using IBM SPSS Statistics 21.

Results

Descriptive Statistics and Assumptions

For theoretical reasons, upon calculating a $Z_r$ value for each observation, the absolute value was taken prior to data processing. The means, medians and standard deviations for the absolute Fisher $Z_r$ values, and number of observations per comparison group are presented in Table 1. Given, the current research proposes that systematic differences in effect size magnitude exist between measures of trial-by-trial vividness ratings and the VVIQ, and no speculation is made about the directionality of the results, the inclusion of negative $Z_r$ values would only serve to underestimate the magnitude of the overall effect size within each comparison groups. However, converting the data to its absolute value caused the data to appear positively skewed. In an effort to accommodate assumptions of normality and homogenize the variance, such that more sensitive parametric hypothesis testing could be utilized, a $2.564\sqrt{Z_r}$ transformation was applied. Levene’s test did not indicate a violation to the equality of variance assumption $F(3, 3528) = 2.605, p = 0.05$, nor did Kolmogorov-Smirnov indicate a violation to the normality assumption,
\[ F(1, 3532) = 0.015, \ p = 0.066. \] The means, medians and standard deviations for the transformed Fisher \( Z_r \) values, and number of observations per comparison group are presented in Table 2.

Table 1

*Descriptive statistics for comparison groups (untransformed absolute \( Z_r \)).*

<table>
<thead>
<tr>
<th>Rating</th>
<th>Measure</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vividness</td>
<td>Behavioural/Cognitive</td>
<td>0.4209</td>
<td>0.3428</td>
<td>0.3422</td>
<td>1760</td>
</tr>
<tr>
<td></td>
<td>Neuroscientific</td>
<td>0.7652</td>
<td>0.7765</td>
<td>0.3331</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.4358</td>
<td>0.3541</td>
<td>0.3489</td>
<td>1840</td>
</tr>
<tr>
<td>VVIQ</td>
<td>Behavioural/Cognitive</td>
<td>0.3246</td>
<td>0.2427</td>
<td>0.2991</td>
<td>1629</td>
</tr>
<tr>
<td></td>
<td>Neuroscientific</td>
<td>0.7960</td>
<td>0.7582</td>
<td>0.4225</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.3421</td>
<td>0.2554</td>
<td>0.3173</td>
<td>1692</td>
</tr>
<tr>
<td>Total</td>
<td>Behavioural/Cognitive</td>
<td>0.3823</td>
<td>0.2941</td>
<td>0.3325</td>
<td>3389</td>
</tr>
<tr>
<td></td>
<td>Neuroscientific</td>
<td>0.7788</td>
<td>0.7753</td>
<td>0.4582</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.3909</td>
<td>0.3067</td>
<td>0.3373</td>
<td>3532</td>
</tr>
</tbody>
</table>

*Non-Parametric Analysis*

Preliminary non-parametric analyses were performed to test the general pattern of results. A Kruskal-Wallis test was conducted to determine whether the median \( Z_r \) values were different between the “Vividness and Behavioural/Cognitive,” “VVIQ and Behavioural/Cognitive,” “Vividness and Neuroscientific,” and “VVIQ and Neuroscientific” groups. The median \( Z_r \) value was significantly different between the four comparison groups, \( \chi^2 (3, N = 3532) = 237, \ p<0.0001 \). A Mann-Whitney test indicated that the median transformed \( Z_r \) value for Vividness (\( \text{Mdn} = 0.6670 \)) was greater than the VVIQ (\( \text{Mdn} = 0.5873 \)), when collapsed across Measure, \( U \).
When collapsed across Rating, a Mann-Whitney test indicated that the median transformed $Z_r$ value for Neuroscientific (Mdn = 0.9055) was greater than Behavioural/Cognitive (Mdn = 0.6204), $U = 95388, p < 0.0001$.

Table 2

*Descriptive statistics for comparison groups (transformed absolute $Z_r$)*

<table>
<thead>
<tr>
<th>Rating</th>
<th>Measure</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vividness</td>
<td>Behavioural/Cognitive</td>
<td>0.6544</td>
<td>0.6587</td>
<td>0.2314</td>
<td>1760</td>
</tr>
<tr>
<td>Neuroscientific</td>
<td></td>
<td>0.8706</td>
<td>0.9061</td>
<td>0.1946</td>
<td>80</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>0.6638</td>
<td>0.6670</td>
<td>0.2341</td>
<td>1840</td>
</tr>
<tr>
<td>VVIQ</td>
<td>Behavioural/Cognitive</td>
<td>0.5818</td>
<td>0.5756</td>
<td>0.2251</td>
<td>1629</td>
</tr>
<tr>
<td>Neuroscientific</td>
<td></td>
<td>0.8699</td>
<td>0.8977</td>
<td>0.2375</td>
<td>63</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>0.5925</td>
<td>0.5873</td>
<td>0.2320</td>
<td>1692</td>
</tr>
<tr>
<td>Total</td>
<td>Behavioural/Cognitive</td>
<td>0.6195</td>
<td>0.6204</td>
<td>0.2313</td>
<td>3389</td>
</tr>
<tr>
<td>Neuroscientific</td>
<td></td>
<td>0.8703</td>
<td>0.9055</td>
<td>0.2138</td>
<td>143</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>0.6297</td>
<td>0.6307</td>
<td>0.2358</td>
<td>3532</td>
</tr>
</tbody>
</table>

*Parametric Analyses*

In an effort to further understand the relationship between the four comparison groups, a 2 (Rating: Vividness, VVIQ) x 2 (Measure: Behavioural/Cognitive, Neuroscientific) ANOVA was performed. There was a significant main effect Measure $F(1, 3528) = 166, p < 0.0001$; however, the main effect of Rating only approached significance $F(1, 3528) = 3.50, p = .061$. The interaction also approached significance $F(1, 3528) = 3.38, p = .066$. Follow up t-tests revealed that for Rating, the mean Vividness $Z_r$ value was statistically greater for Neuroscientific
measures ($M = 0.8706, SD = 0.1946$) than Behavioural/Cognitive measures ($M = 0.6544, SD = 0.2314$), $t(89.472) = -9.631, p<0.0001$ (equal variance not assumed). This pattern of results was also the case for the VVIQ, where the average $Z_r$ value was statistically greater for Neuroscientific measures ($M = 0.8699, SD = 0.2375$) than Behavioural/Cognitive measures ($M = 0.5818, SD = 0.2251$), $t(66.38) = -9.466, p<0.0001$ (equal variance not assumed). For Measure, the mean Behavioural/Cognitive $Z_r$ value was greater for Vividness ($M = 0.6544, SD = 0.2314$) than the VVIQ ($M = 0.5818, SD = 0.2251$), $t(3387) = 9.249, p<0.0001$ (equal variance assumed). However, the mean Neuroscientific $Z_r$ value was not significantly different between Vividness ($M = 0.8706, SD = 0.1946$) and VVIQ ($M = 0.8699, SD = 0.2375$), $t(141) = 0.18, p < 0.986$ (equal variance assumed).

The notion that the main effect of Rating was not significant, when collapsed across Measure, is at odds with the preliminary non-parametric analysis which showed this effect to be highly significant. As evidenced by Table 2, given the extensive sample size differences between Rating (N = 3389) and Measure (N = 143), problems of intercorrelation may have been introduced on account of the default use of unweighted means in SPSS. As a subsidiary analysis, a 2 (Rating: Vividness, VVIQ) x 2 (Measure: Behavioural/Cognitive, Neuroscientific) ANOVA with unequal sample sizes was performed, where the interaction was considered after the main effects. The overall analysis including the interaction term was significant $F(3, 3528) = 84.0, p<0.0001$, as were the main effects of Rating $t(3530) = -9.08, p<0.0001$, and Measure $t(3530) = 12.7, p<0.0001$. These data are illustrated in Figure 2 and Figure 3, respectively. However, the interaction between these variables was not significant $t(3530) = 1.84, p = .066$. 
Figure 2. Mean Transformed $Z_r$ for Rating collapsed across Measure. Error bars indicate 95% confidence interval.
Figure 3. Mean Transformed Zr for Measure collapsed across Rating. Error bars indicate 95% confidence interval.

Table 3

Descriptive statistics for comparison groups (transformed absolute Zr, outliers removed)

<table>
<thead>
<tr>
<th>Rating</th>
<th>Measure</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vividness</td>
<td>Behavioural/Cognitive</td>
<td>0.6509</td>
<td>0.6587</td>
<td>0.2267</td>
<td>1751</td>
</tr>
<tr>
<td></td>
<td>Neuroscientific</td>
<td>0.8706</td>
<td>0.9061</td>
<td>0.1946</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.6605</td>
<td>0.6670</td>
<td>0.2298</td>
<td>1831</td>
</tr>
<tr>
<td>VVIQ</td>
<td>Behavioural/Cognitive</td>
<td>0.5803</td>
<td>0.5752</td>
<td>0.2225</td>
<td>1626</td>
</tr>
<tr>
<td></td>
<td>Neuroscientific</td>
<td>0.8699</td>
<td>0.8977</td>
<td>0.2375</td>
<td>63</td>
</tr>
</tbody>
</table>
A total of 12 extreme observations, which exceeded the value of 1.27 were removed from the dataset, at which point Levene’s test did not indicate a violation to the equality of variance assumption $F(3, 3516) = 2.428, p = 0.064$, and normality only marginally exceeded the standard 0.05 cut-off as per Kolmogorov-Smirnov, $F(1, 3520) = 0.016, p = 0.047$. The means, medians and standard deviations for the transformed Fisher $Z$ values, and number of observations per comparison group are presented in Table 3. Non-parametric and parametric analyses were performed in the same manner as before.

**Non-Parametric Analysis (Outliers Removed)**

A Kruskal-Wallis test was conducted to determine whether the median $Z$ values were different between the “Vividness and Behavioural/Cognitive,” “VVIQ and Behavioural/Cognitive,” “Vividness and Neuroscientific,” and “VVIQ and Neuroscientific” groups. The median $Z$ value was significantly different between the four comparison groups, $\chi^2(3, N = 3520) = 237, p<0.0001$. A Mann-Whitney test indicated that the median transformed $Z$ value for Vividness ($Mdn = 0.6670$) was greater than the VVIQ ($Mdn = 0.5863$), when collapsed across Measure, $U = 1269118, p<0.0001$. A Mann-Whitney test indicated that the median transformed $Z$ value for Neuroscientific ($Mdn = 0.9055$) was greater than Behavioural/Cognitive ($Mdn = 0.6194$), when collapsed across Rating, $U = 93672, p<0.0001$.

**Parametric Analyses (Outliers Removed)**
A 2 (Rating: Vividness, VVIQ) x 2 (Measure: Behavioural/Cognitive, Neuroscientific) ANOVA was performed. There was a significant main effect Measure $F(1, 3516) = 174, p<0.0001$; however, the main effect of Rating only approached significance $F(1, 3516) = 3.42, p = .065$. The interaction also approached significance $F(1, 3516) = 3.29, p = .070$. Follow up t-tests revealed that for Rating, the mean Vividness $Z_r$ value was statistically greater for Neuroscientific measures ($M = 0.8706, SD = 0.1946$) than Behavioural/Cognitive measures ($M = 0.6509, SD = 0.2267$), $t(89.087) = -9.799, p<0.0001$ (equal variance not assumed). This was pattern of results was also the case for the VVIQ, where the average $Z_r$ value was statistically greater for Neuroscientific measures ($M = 0.8699, SD = 0.2375$) than Behavioural/Cognitive measures ($M = 0.5803, SD = 0.2225$), $t(66.287) = -9.519, p<0.0001$ (equal variance not assumed). For Measure, the mean Behavioural/Cognitive $Z_r$ value was greater for Vividness ($M = 0.6509, SD = 0.2267$) than the VVIQ ($M = 0.5803, SD = 0.2225$), $t(3375) = 9.125, p<0.0001$ (equal variance assumed). However, the mean Neuroscientific $Z_r$ value was not significantly different between Vividness ($M = 0.8706, SD = 0.1946$) and VVIQ ($M = 0.8699, SD = 0.2375$), $t(141) = 0.18, p < 0.986$ (equal variance assumed).

Again, given the differences in sample size between Rating ($N = 3377$) and Measure ($N = 143$), a 2 (Rating: Vividness, VVIQ) x 2 (Measure: Behavioural/Cognitive, Neuroscientific) ANOVA with unequal sample sizes was performed, where the interaction was considered after the main effects. The overall analysis was significant $F(3, 3516) = 86.0, p<0.0001$, as were the main effects of Rating $t(3518) = -8.96, p = .000$, and Measure $t(3518) = 13.1, p<0.0001$. These data are illustrated in Figure 4 and Figure 5, respectively. However, the interaction between these variables was not $t(3518) = 1.81, p = .070$. 
Figure 4. Mean Transformed $Z_r$ for Rating (Vividness and VVIQ) collapsed across Measure upon removing outliers. Error bars indicate 95% confidence interval.
Unique Observations

Unique observations correlating Vividness with the VVIQ were considered independently of the comparison groups. Although the absolute value for these data was computed, for the purposes of interpretability, these values were not subjected to the same normalization transformations. The mean $Z_t$ value for the 47 observations correlating Vividness with the VVIQ was 0.385 (SD = 0.3453), which ranged from 0.01 to 1.28. These $Z_t$ values correspond to an average $r$ of 0.367 (SD = 0.332) ranging from 0.01 to 0.856.
Discussion

Interpretation of Results

The present research synthesized, analyzed and evaluated a robust subset of literature pertaining to trial-by-trial vividness ratings and the VVIQ in an effort to understand, and empirically validate the construct of vividness as a critical aspect to the mental imagery process. Using meta-analytic methodology, trial-by-trial ratings of vividness and the VVIQ were compared against any and all available behavioural/cognitive and neuroscientific measures, such that an average unstandardized effect size ($r$) could be calculated, and contrasted. In order to validate the vividness construct, the present research assumed that the effect size magnitude for each comparison group would be indicative of the propensity for that variable to measure that which it is presumed to measure (construct validity), and the extent to which it is capable of measuring all facets of it (content validity). Because generalized questionnaires concerning individual differences in the subjective experience of vividness insufficiently resolve critical variability in the mental imagery process, it was hypothesized that trial-by-trial ratings of vividness would result in an average effect size magnitude which was greater than that of the VVIQ. One may consider this hypothesis a test of validation of content, examining the extent to which trial-by-trial rating of vividness and the VVIQ measure all encompassing aspects of the vividness construct. As evidenced by Table 2 and Table 3, and as illustrated in Figure 2 and Figure 4, the average $Z_r$ value for ratings of vividness was statistically greater than that of the VVIQ, which supports the hypothesis that trial-by-trial ratings of vividness demonstrate construct and content validity which is greater than that of the VVIQ, when collapsed across both behavioural/cognitive and neuroscientific measures.
Given the phenomenological experience of vividness may be representative of variability in brain functioning more directly than behavioural and cognitive measures to which vividness is typically correlated, it was reasoned that neuroscientific measures would sufficiently resolve vividness with a greater propensity than behavioural and or cognitive measures. Although neuroscientific measures may inadvertently measure these correlated processes as well, the average effect size within the neuroscientific domain should demarcate the extent to which vividness can be measured, as it is thought to be an exclusively neurological process. For example, if a variety of independent processes were measured via fMRI, and all were found to be moderately related to vividness, subtractive techniques could reduce the construct to its isolated neuroanatomical parts. Here, it is argued that this option is not available for behavioural and or cognitive measures. Accordingly, the average effect size magnitude for neuroscientific measures was hypothesized to be greater than that of behavioural and or cognitive measures. One may consider this hypothesis as a test of validation of construct, examining the degree to which behavioural/cognitive and neuroscientific measure that which they purport to measure. Also evidenced by Table 2 and Table 3, and illustrated in Figure 3 and Figure 5, is the systematic difference between the measures against which vividness and the VVIQ are compared, when collapsed across trial-by-trial ratings and the VVIQ. This finding lends support to this contention that neuroscientific measures resolve the phenomenological experience of vividness with greater propensity than behavioural and or cognitive ones. With respect to Table 1 specifically, the untransformed absolute effect size magnitude for behavioural/cognitive measures corresponds to a $Z_r$ value of 0.3823 ($r = 0.365$), and neuroscientific measures to a $Z_r$ value of 0.7788 ($r = 0.652$). According to Cohen (1988), the difference between these two $r$-values parallels the differences between a medium and strong effect size, respectively. Furthermore, the notion that significantly
greater effect sizes accompany neuroscientific measures may suggest that vividness is an inherent aspect of the psychological experience, one which is only observable through neuroimaging. In this sense, behavioural and cognitive effect sizes may be the unintended, yet direct measure of secondary process to which vividness is correlated.

In an effort to further explore the meaning of the results all comparison groups were independently contrasted. When compared within the context of vividness ratings, the average effect size magnitude for neuroscientific measures was statistically greater than behavioural/cognitive measures. This trend was consistent for the VVIQ as well. Given the average effect size magnitude for neuroscientific measures significantly exceeded the magnitude of behavioural/cognitive ones for both vividness ratings and the VVIQ, the stability of this observation can be interpreted as convergent validity. The convergent validity offered by vividness ratings and the VVIQ bolsters the previous contention, wherein neuroscientific measures offer more construct validity than behavioural/cognitive ones, an observation of which is further corroborated by the absence of a significant interaction. When vividness ratings and the VVIQ were compared within the context of behavioural/cognitive measures, the average standard effect size for vividness ratings exceeded that of the VVIQ. Again, this observation can be interpreted as a validation of content, where trial-by-trial ratings of vividness offer a more encompassing estimate of all aspects of the vividness construct. However, no statistically significant differences were found between trial-by-trial ratings of vividness and the VVIQ within the context of neuroscientific measures. Despite this non-significant finding, the present research may provide preliminary evidence demonstrating the precision and reliability of trial-by-trial ratings of vividness in tandem with neuroscientific measures is a most effective means of resolving the vividness construct. For example, because a significant effect for Measure and a
significant effect for Rating were found, the absence of a significant interaction between Measure and Rating may suggest the combination of neuroscientific measures and trial-by-trial ratings provide a most reliable measure. According to this interpretation, the central tendency of the neuroscientific comparison groups may be biased due to a relatively small sample size, or perhaps another mediating factor. This view can be challenged, however, as the interaction approached significance ($p = 0.066$), but the neuroscientific comparison did not ($p = 0.986$).

Unique observations concerning the nature of the relationship between trial-by-trial ratings of vividness and the VVIQ did not demonstrate strong agreeability ($Z_r = 0.385; r = 0.367$). Although trial-by-trial ratings and the VVIQ purportedly measure vividness, the lack of a strong correlation between the rating systems implies they do not resolve the vividness construct to the same degree. Instead, it would suggest that these two measurements of vividness resolve some underlying property of mental imagery, but differ widely in the content of their measurement. For example, trial-by-trial ratings and the VVIQ may be an adequate measure an individual’s propensity to form mental images more generally (trait), but trial-by-trial ratings may also resolve situational variability within an individual (state). Such situational variability may be consistent with flashbulb memories, pharmaceutical interventions, hallucinations, or dreams. This conclusion is consistent with the data relating to the behavioural/cognitive comparison groups, wherein the magnitude of the effect sizes relating to trial-by-trial ratings of vividness is larger than that of the VVIQ. It is likely that trial-by-trial ratings of vividness resolve situational differences in the mental imagery process than the VVIQ is capable of. However, as to why neuroscientific measures did not abide by the same trend as the behavioural/cognitive ones in this respect is a point that is to be discussed.

*Appropriateness of the Meta-Analytic Model*
Given the present investigation is contingent on the meta-analytic methodology, limitations and disadvantages inherent in meta-analysis are likewise concerns limiting the interpretation of the present results. Rosenthal (1991) identifies six such criticisms, which include 1) sampling bias and the file drawer problem, 2) information loss, 3) problems of heterogeneity, 4) problems of independence, 5) exaggeration of significance levels, and 6) the practical importance of the estimated effect size. The present investigation considered said criticisms both a priori and posteriori, as to minimize the influence of problems inherent in meta-analysis, and increase the general strength of the key findings. Sampling bias and the file drawer problem (1) criticisms posit that the sample of retrieved studies in meta-analysis does not truly represent the population of research conducted on the subject. For example, the publication process may not favour experimental outcomes wherein the null hypothesis is accepted, such that non-significant findings are not communicated, nor are they available for later retrieval. Subsequently, some critics contend that meta-analysis over-inflates the estimated effect size of the variable in question, and some even contend that meta-analysis in nothing more than the sum total of Type-1 error.

Rosenthal and Rubin (1988) provide strong evidence that the file drawer problem and publication bias may not be as prevalent in meta-analysis research as opponents would contend. These researchers compared to groups of studies on a particular subject, one of which could be completely retrieved as they belonged to a single laboratory (complete), and the other of which could not (incomplete). In addition, these researchers considered three publication outcomes (published immediately, eventually published (within 18 years), or never published). Although complete retrieval was indeed found to result in less significant outcomes than incomplete retrieval, and unpublished studies were found to results in less significant outcomes than
immediately published studies, studies which were eventually published showed less significant outcomes than unpublished studies. These results suggest that meta-analysis may reduce over-inflation, and perhaps even result in a more conservative estimate of effect sizes than the true values, provided a long enough period of time has transpired. Because the present analysis includes papers ranging from 1966 to 2012, the data reported in Table 1 specifically, may be relatively unbiased or at least conservative estimates of the actual effect sizes they represent with respect to publication bias. With that being said, neuroscientific measures typically displayed a much narrower date range of publication, such that these data may be inflated. With the exception of three entries dating from 1983 to 1989, neurophysiologic data ranged from 1995 to 2012, which may suggest a systematic bias exists for this comparison group given this range falls within the critical 18 year window.

It is worth noting that most neuroscientific entries concerning fMRI data were parcelled into an even narrower window. This point may explain the significantly larger effect sizes for neuroscientific data, but also explain the lack of a significant contrast between vividness ratings and the VVIQ along the neuroscientific dimension. Nevertheless, although irretrievability may have influenced the practical interpretation of the present results, publication bias is not thought to have influenced the results more generally. In addition, extending the versatility of Rosenthal’s (1979) fail-safe N (Nfs), concerning sampling bias in meta-analysis for probability levels, Orwin (1983) introduced a fail-safe N for the effect size d (See Equation 2), transcribed for effect size r (See Equation 3). Fail-safe N refers to the quantity of statistical outcomes reporting a zero effect size (dfs = 0), which would be required to reduce the strength of an effect to a certain level of intolerance (d_c). For example, Cohen (1988) defined d values of 0.2, 0.5, and 0.8 as small, medium and large effect sizes, respectively. As per Orwin’s (1983) example, Smith,
Glass, and Miller (1980) reported an average effect size $d = 0.85 (d_0)$ for psychotherapy outcomes, as per a meta-analysis involving 1766 ($N_0$) effect size observations. Accordingly, the number of unreported effect sizes equalling zero, which would be required to reduce the estimated effect size from $d = 0.85$ to $d = 0.2$ would be approximately 5740. The rather large number for $N_{fs}$ seems like an unlikely number of null effect sizes to exist unpublished.

$$N_{fs} = \frac{N_0 (d_0 - d_c)}{d_c - d_{fs}}$$

Equation 2: Calculation of Fail-safe N for Cohen’s $d$. The quantity of unpublished effect sizes required ($N_{fs}$) reporting $d = 0 (d_{fs})$ to reduce the observed effect size ($d_0$), averaged over the number of observations ($N_0$) to some critical value ($d_c$).

$$d = \frac{4r^2}{\sqrt{1 - r^2}}$$

Equation 3: Calculation of Cohen’s $d$ from effect size $r$ for Equation 2.

With respect to the present investigation, from Table 1, 3532 effect sizes ($N_0$) were observed in total for the comparison groups, with an average $Z_r = 0.3909 (r = 0.372 (d_0 = 0.802))$. Assuming $d_{fs} = 0$, the number of unpublished effect sizes ($N_{fs}$) required to reduce the observed average effect size to a value of $d_c = 0.2 (r = 0.1$ (small effect)) would be 10,631. The practical purpose of this exercise is to demonstrate that the file drawer problem may exist to a certain degree in meta-analysis, yet the contribution of this limitation is seemingly negligible as the number of observations increases, as it would seem most unlikely that such a large quantity of effect sizes equalling zero would exist unpublished. To summarize, with the exception of neuroscientific measures, there appears to be no immediate reason to conclude that publication
bias should influence the comparison groups in any systematic way. In addition, the overall
effect size of medium strength \((r = 0.372)\) would appear to be fairly reliable, given the large
number of unpublished effect sizes reporting an \(r\)-value of zero which would be required to
reduce the average \(r\)-value to less than 0.1.

Rosenthal (1991) referred to information loss (2) as an inherent limitation of meta-
analysis, yet acknowledges that this limitation is perhaps inherent in data analysis more
generally. The contention is one which criticizes the overemphasis on single values, and glossing
over details, both of which may exist in the present data to a certain degree. For example, entire
publications were at times summarized by a single statistic, and each observation was recorded
as the mean of some statistical outcome, wherein some variability was present, yet not taken into
consideration in the present investigation. Perhaps a more concerning criticism in the present
investigation was the overlooking of negative values. Although any and all statistical outcomes
were recorded, all practical interpretability of the mean effect size for each comparison group
was partially compromised upon transforming each observation into its absolute value, and
virtually abandoned upon transforming these data to accommodate normality and equal variance
assumptions. Nevertheless, the present investigation questioned the magnitude, not the
directionality of the average effect size for each comparison group in relation to each other.
Given each comparison group was treated in the same way, one may consider the current
research as a trade-off between interpretability and generalizability. Such a consideration is also
applicable to problems of heterogeneity (3), and to a lesser degree, problems of independence
(4). Although the present research may compromise the extent to which future researchers may
interpret the average effect size for practical purposes, the general findings concerning the
comparability of the four comparison groups is thought to be fairly robust.
Problems of heterogeneity (3), or “combining apples and oranges” (Rosenthal & DiMatteo, 2001; Glass, 1978) is a criticism acknowledging the subtle and sometimes obvious differences in research methodologies between the summated measures in meta-analysis research. For example, problems concerning heterogeneity of method include differences in the variable against which vividness ratings or the VVIQ are measured, such as reaction times, gender, emotionality, and event-related potentials, etc. Furthermore, demand characteristics, sample demographics, and procedural nuances may account for large variations in the strength of each effect. Problems concerning heterogeneity of quality refer to the source from whence the data were extracted, such as journal quality. However, these two major criticisms encompassing problems of heterogeneity, and to a lesser degree, problems of independence (4), were reconciled given one assumption. If the sample effect sizes represent a sufficiently large subset of the population, and heterogeneity is inherently variable in the measurement, then the sample mean will be normally distributed with mean $\mu$ and standard deviation $\frac{\mu}{\sqrt{n}}$. That is to say, the error affiliated with the measurement should be randomly distributed about the mean, which should be normally distributed within each comparison group. Although variation may exist in the measures against which vividness ratings and the VVIQ are compared, it is thought this bias should be negligible.

Although problems of independence (4) were assumed to be captured by the aforementioned assumption concerning problems of heterogeneity, notable exceptions in the dataset may be challenged by proponents of this criticism. Problems of independence generally concern responses within studies and studies within sets of studies. For example, the results from one group of researchers may be more similar to other works by the same researchers, than they are to works by other researchers. In addition, variability may exist in the number of statistics
reported by a group of researchers, which is a particular issue in the present investigation, where one study reported a single effect falling under the purview of the research question (Borkovec & Krogh Sides, 1979), while another reported 116 (Kunzendorf, 1985-1986). Because commentators may regard independence as a shortcoming of meta-analysis, and the mean effect sizes reported in each comparison group may be variably influenced by such problems, this issue represent an exceptional limitation in the present research. Although Rosenthal (1969) demonstrated that problems of independence may be generally negligible in meta-analysis with respect to probability-based meta-analysis, it is not known whether this contention is applicable to meta-analyses concerning effect sizes.

Both exaggeration of significance levels (5) and the practical importance of the estimated effect size (6) are of little concern to the present analysis. The former criticism (5) cautions against the truncation of significance levels affiliated with rounding Z-scores and p-values, and also cautions against the inclusion of too many studies. However, the tenability of these arguments is only true concerning probability-based meta-analyses, not effect size-based ones. Although relevant to the unique observations correlating vividness ratings with the VVIQ, the practical importance of the estimated effect size (6) is an argument established to question the meaning and interpretability of r and r^2 in social science research. In the present investigation, unique observations were recorded and calculated independently such that the relationship between vividness ratings and the VVIQ could be clearly understood. Whether these values are to be used to promote practical purposes is beyond the scope of the present analysis. Instead, these values were to demonstrate that the convergent validity between ratings of vividness and the VVIQ is weak to moderate. Despite face validity, content validity seems to be disparate
between these measures, an observation corroborated by the relative small $r$ value between these measures.

Theoretical Considerations and Future Directions

A clinical characteristic of PTSD is the vivid and involuntary retrieval of a traumatic memory, which typically occurs through the visual modality, and accompanied by negative affect such as fear, helplessness, guilt, anger, shame and horror (Hackmann, 2011). In this sense, PTSD consists of two components, a cognitive (vividness) component, and a (physiological) emotionality component. Given its concomitant relationship with emotionality (Shin, Rauch, & Pitman, 2006; Soderlund, Moscovitch, Kumar, Mandic, & Levine, 2012), the conceptualization of vividness as an empirical research prerogative may promote advances in mental health theory, research and practice. Evermore relevant to EMDR, the validation of the vividness construct may facilitate a resolution of neurobiological-based and cognitive-based theoretical frameworks, the disparity of which is of great interest to researchers (Bergmann, 2010; van den Hout & Engelhard, 2012). Many contemporary neurobiological theories attribute the effectiveness of EMDR to the regulatory role of the ACC in cognitive and affect processes, which may be reasonable given brain morphology (Devinsky, Morrell, & Vogt, 1995), and the presumed neurobiological underpinnings of PTSD (Richardson, Strange, & Dolan, 2004; for a review see Phelps, 2004). However, the vividness construct itself may be trivialized (Corrigan, 2002; Kaye, 2007; Harper, Rasolkhani-Kalhorn, Drozd, 2009). Moreover, evidence for this contention is available from the present research. Despite significantly larger effect size magnitudes, experimental outcomes pertaining to the explicit measurement of vividness within the realm of neuroscientific represented approximately 4% of the current database (Vividness ratings and VVIQ combined; see Table 1). Since recent clinical evidence suggests changes in vividness
precede those of emotionality (Smeets et al., 2012), and neuroimaging techniques provide a feasible means to study the subjective experience of vividness (Gonsalves & Paller, 2002; Cui, Jetter, Yang, Montague & Eagleman, 2007), theoretical revisions may be an imminent realization for EMDR neurobiological theorists. Although the ACC may mediate the cognitive and affective processes affiliated with EMDR, it is perhaps practical to assume that these processes are synonymous with the measures of vividness and emotionality, respectively.

Under such a dichotomous conceptualization, neurobiological theories may accommodate a growing corpus of research in the EMDR field, yet require only minimal modification to their premises. For example, D’Angiulli, Runge, Faulkner, Zakizadeh, Chan, and Morcos (2013) recently proposed a neurobiological theory concerning the construct of mental imagery vividness. Predicated on multi-trace memory theory (MMT; Moscovitch et al., 2005), these authors assert that vividness ratings represent a direct translation of residual top-down sensory traces available in autobiographical memory, wherein each sensory trace is a distributed pattern of cortical activation, and unique to a specific sensory stimulus (Hintzman, 1976). Upon encoding a memory, the hippocampus is thought to index sensory traces for later retrieval (Ryan et al., 2001); however, the influence of this structure becomes less influential as sensory traces are consolidated into existing cortical networks through repetitious presentations (Takashima et al., 2009). Mental images are integrated into a coherent whole through the cuneus, precuneus and occipital regions (Svoboda, McKinnon, & Levine, 2006; Cabeza & St. Jacques, 2007). The phenomenological experience of vividness is thought to be proportional to the associative strength of the sensory trace, and the speed at which the mental image forms.

Within the context of EMDR, if vividness is the phenomenological representation of sensory traces throughout the cortex, the ACC may mediate the subjective experience of
vividness—and presumably emotionality—through the cognitive and affective subdivisions, respectively. Neuronal pathways from the cognitive subdivision to the PCC may contribute to the retrieval of autobiographical memories (Maddock, Garrett, & Buonocore, 2001) en route to the precuneus, wherein the phenomenological experience of conscious information processing and mental imagery occur (Vogt & Laureys, 2005; Fletcher et al., 1995; Cavanna & Trimble, 2006). Projections transmitting eye-movement information from the thalamus unto the cognitive subdivision (specifically, the PCC; Vogt, Finch, & Olsen, 1992) may significantly interfere with the process of autobiographical memory recall, and mental imagery formation. Although Gunter and Bodner (2008) posit eye movements consume attentional resources otherwise utilized in vivid imagining and emotional processing, the present argument assumes that strictly vivid imagining is interfered with. Although changes in vividness are predicted to precede those of emotionality (Smeets et al., 2012), eye-movements are not predicted to immediately interfere with emotionality. Instead, the present theory supposes that changes emotionality are not immediate, and occur as a direct results of altered vividness, perhaps through a transitory process involving memory consolidation. In PTSD, the recurrent intrusion of an imminent threat affords a distorted appraisal of the traumatic stimulus, one which is disconnected from autobiographical memory (poorly consolidated), yet robustly associative (Ehlers & Clark, 2000; Conway & Pleydell-Pearce, 2000). In this manner, EMDR acts to facilitate consolidation, and impede the association of memories for the traumatic event, which are speculated to operate through a transitory, iterative process.

One may consider the EMDR procedure as a series of iterative steps, whereby the continual interference of thalamic projections unto the PCC progressively decrements the subjective experience of vividness. Every iterative step coincides with stepwise decrease in the
validness of the mental image, and a corresponding consolidation of the original sensory trace. The observed decrease in emotionality at any given time in the EMDR procedure is subsequent of the memory consolidation which occurred at the previous iterative step, which itself occurred as a direct result of altered vividness. Furthermore, this postulation can be computationally modelled. Because the “vivid-is-fast” relationship asserts vivid images tend to be rendered more quickly than images which are not (D’Angiulli & Reeves, 2002), systematic reductions in vividness would be expected to be met with increased image generation times. If one assumes EMDR operates as an iterative process, where each iterative step is marked by a stepwise decrease in vividness, and occurs within a certain period of time, then successive iterations would be predicted to take longer and longer. Vividness—and shortly thereafter, emotionality—would be predicted to asymptotically decrease according to an inverse power function. The inclusion of reaction times for ratings of vividness and emotionality in future experiments employing an EMDR paradigm may provide critically informative data concerning the mental imagery process and perhaps the memory consolidation process as well. In addition, if trial-by-trial ratings of vividness can be correlated with neurophysiologic markers, such as event-related amplitude and latency, within the context of an EMDR paradigm, the resultant data may advance theoretical understanding of the cognitive and neurobiological substrates underpinning mental illness. Although self-reports of vividness have been sufficiently correlated with neuroscientific measurement (Gonsalves & Paller, 2002; Cui, Jetter, Yang, Montague & Eagleman, 2007), and EEG has been employed in EMDR research (Harper, Rasolkhani-Kalhorn, Drozd, 2009), the vividness construct has not been systematically investigated with neuroscientific measures within the context of EMDR. The results of the present investigation suggest that vividness may
represent a viable empirical construct, the measurement of which may support a long awaited 
resolve in EMDR neurobiological theorization.

Conclusions

Despite synonymous operational definitions, the results of the present meta-analysis suggest that 
trial-by-trial ratings of vividness and the VVIQ differ greatly in the content and the extent of 
what they purport to measure. It was demonstrated that against any and all available behavioural 
and cognitive measures, the average effect size magnitude was greater for ratings of vividness on 
a trial-by-trial basis than the VVIQ. This trend was also observed for the average effect size 
magnitude when behavioural/cognitive and neurophysiologic measures were collapsed. Because 
effects sizes for trial-by-trial ratings were generally greater than those of the VVIQ, this finding 
can be interpreted as support for content validity in favour of trial-by-trial ratings of vividness 
over the VVIQ, and perhaps global assessments of imaging ability more generally. Although 
trial-by-trial ratings and the VVIQ may resolve an individual’s ability to experience vivid mental 
imagery, it is thought that trial-by-trial ratings sufficiently resolve critical variability in the 
content of the imagery experience, the like of which the VVIQ cannot. For neurophysiologic 
measures in isolation, no statistical difference in the average effect size magnitude was observed 
between trial-by-trial ratings of vividness and the VVIQ. However, when collapsed across trial-
by-trial ratings of vividness and VVIQ, the average effect size magnitude was greater for 
neurophysiologic measures than behavioural/cognitive ones. This finding can be interpreted as 
general support for the vividness construct. Taken together, neuroscientific measures, in 
conjunction with trial-by-trial ratings of mental imagery, may provide a more precise and reliable 
measure of vividness in clinical and research settings. Lastly, despite face validity, unique 
observations correlating trial-by-trial rating of vividness with the VVIQ showed a modest to
medium strength relationship. Given the statistically greater effect size magnitude affiliated with trial-by-trial ratings of vividness, this disparity should prompt serious consideration for researchers interested in the study of mental imagery. Trial-by-trial ratings of vividness may resolve critical aspects of the mental imagery process, the like of which the VVIQ is unable to. The present research has sought to establish a methodological precedent for future researchers interested in the study of mental imagery. The practicality of the present investigation is one which serves to establish the empirical construct of vividness, such that mental imagery may contribute to mental health science, research, theory, and practice. Although a theoretical framework concerning EMDR was herein delineated as a paradigmatic example, it is anticipated that the results of the present research will be more widely applicable.
References


Appendix A

The VVIQ (Marks, 1973). Subjects are asked to rate the vividness of the mental image on a scale from 1 (‘Perfectly clear and vivid as normal vision’) to 5 (‘No image at all, you only “know” that you are thinking of the object’).

Items contained in the Vividness of Visual Imagery Questionnaire

For items 1–4, think of some relative or friend whom you frequently see (but who is not with you at present) and consider carefully the picture that comes before your mind’s eye.

Item
1. The exact contour of face, head, shoulders and body.
2. Characteristic poses of head, attitudes of body, etc.
3. The precise carriage, length of step, etc., in walking.
4. The different colours worn in some familiar clothes.

Visualize a rising sun. Consider carefully the picture that comes before your mind’s eye.

Item
5. The sun is rising above the horizon into a hazy sky.
6. The sky clears and surrounds the sun with blueness.
7. Clouds. A storm blows up, with flashes of lightning.
8. A rainbow appears.

Think of the front of a shop which you often go to. Consider the picture that comes before your mind’s eye.

Item
9. The overall appearance of the shop from the opposite side of the road.
10. A window display including colours, shapes and details of individual items for sale.
11. You are near the entrance. The colour, shape and details of the door.
12. You enter the shop and go to the counter. The counter assistant serves you. Money changes hands.

Finally, think of a country scene which involves trees, mountains and a lake. Consider the picture that comes before your mind’s eye.

Item
13. The contours of the landscape.
14. The colour and shape of the trees.
15. The colour and shape of the lake.
Appendix B

Formulae for the Conversion of Test Statistics to Standard Effect Size r (Friedman, 1982).

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$</td>
<td>$r = \frac{t^2}{\sqrt{t^2 + df}}$</td>
</tr>
<tr>
<td>$z$</td>
<td>$r = \frac{z^2}{\sqrt{z^2 + N}}$</td>
</tr>
<tr>
<td>$F(df_n = 1)$</td>
<td>$r = \frac{F}{\sqrt{F + dfd}}$</td>
</tr>
<tr>
<td>$F(df_n &gt; 1)$</td>
<td>$r = \frac{dfnF}{\sqrt{dfnF + dfd}}$</td>
</tr>
<tr>
<td>$\chi^2(df_n = 1)$</td>
<td>$r = \frac{\chi^2}{\sqrt{N}}$</td>
</tr>
<tr>
<td>$\chi^2(df_n &gt; 1)$</td>
<td>$r = \frac{\chi^2}{\sqrt{\chi^2 + N}}$</td>
</tr>
</tbody>
</table>

Note: $df_n =$ degrees of freedom (numerator); $dfd =$ degrees of freedom (denominator)
Appendix C

Complete Meta-Analysis Reference List. References are listed according to database from which they were taken.

**Embase**


VALIDATING PHENOMENOLOGICAL ASPECTS


General, 112(4), 662-685.


Holmes, E. A., Deeprose, C., Fairburn, C. G., Wallace-Hadrill, S., Bonsall, M. B., Geddes, J. R.,


Pearson, D. G., Ross, F. D., & Webster, V. L. (2012). The importance of context: evidence that


**Web of Knowledge**


**PsycINFO**


Begg, I. (1988). What does the vividness of an image tell us about the value of imagery?. *Journal of Mental Imagery. 12*(3-4), 45-56


mental imagery-new york-international imagery association, 21, 73-82.


Ewan, L., Smith, N. C., & Holmes, P. S. (2010). Disruption to imagery perspective following stroke: implications for kinesthesis and rehabilitation. *Journal of Mental*


George, L. (1981). A survey of research into the relationships between imagery and psi. *Journal
of Parapsychology. 45(2), 121-146.


Green, M. C. (2004). Transportation into narrative worlds: The role of prior knowledge and perceived realism. *Discourse Processes*, 38(2), 247-266.


Hafner, R. J. (1978). Catharsis during prolonged exposure for snake phobia: An agent of


Heremans, E., Feys, P., Nieuwboer, A., Vercruysse, S., Vandenberghe, W., Sharma, N., & Helsen,


Nixon, R. D., Cain, N., Nehmy, T., & Seymour, M. (2009). The influence of thought suppression and cognitive load on intrusions and memory processes following an analogue...


Price-Williams, D., & Gaines, R. (1994). The Dreamtime and Dreams of Northern Australian


Rossi, A. M., Sturrock, J. B., & Solomon, P. (1963). Suggestion effects on reported imagery in


Sadler, P., & Woody, E. Z. (2006). Does the more vivid imagery of high hypnotizables depend on


Sheehan, P. W. (1971). Individual differences in vividness of imagery and the function of


Weerts, T. C., & Lang, P. J. (1978). Psychophysiology of fear imagery: Differences between


dynamic imagery. *British journal of Psychology, 86*(2), 283-300.


Neuroscience, 23(10), 3146-3152.


