Footnotes

[1] IAP Reference database described at: http://tinyurl.com/dcvrdm

[2] CHC listserv address: http://groups.yahoo.com/group/IAPCHC/; NASP listerv address: http://groups.yahoo.com/group/NASP-Listserv/

[3] In our introduction we argued that a limitation of prior CHC COG-ACH relations synthesis was the inclusion of studies that included a narrow range of CHC IVs (specification error). We could similarly be asked why we included investigations (viz., Ganci, 2004; Keith, 1999; Vanderwood et al., 2002) that did not include measures of the major seven broad (Gf, Gc, Glr, Gsm, Ga, Gv, Gs) domains. We included these studies as, in each instance, the omission of one or more domains was an apriori decision articulated by the researchers based on the failure to find a significant relation for the excluded broad domain (e.g., Gv) in the researchers review of the then extant research literature.

[4] Only school-age samples were included in the review. A number of research investigations reported results for young adult to older adult samples, but collectively the number of adulthood samples was too sparse to allow for meaningful synthesis across studies.

[5] Classification of IV's as g (general; stratum III), broad (stratum II), or narrow (stratum I) was based on the current authors analysis and not necessary that reported by the original researchers. For example, it was a common practice in early CHC research literature to refer to certain measures or factors as reflecting broad abilities (e.g., WJ-R Ga cluster), when in retrospect, later CHC literature recognized that these measures or factors represented narrow CHC abilities (e.g., WJ-R Ga cluster was subsequently determined to be measuring the narrow Ga ability of phonetic coding--PC; see Flanagan & McGrew, 1998 and McGrew &Woodcock, 2001). Another example of IV stratum reclassification was our reclassification of Hale et al.'s (2006) Glr IV as measuring the narrow Glr ability of free recall memory (M6) since this was a single test measure from the DAS (viz., Recall of Objects) and did not measure two or more qualitatively different narrow CHC abilities within the respective broad domain.

[6] See Keith (2006) for definition and differences between manifest (MV) and latent variables (LV).

[7] Samples described by grade level (versus age) were reclassified by age using the following general scheme (6-8 years = grades k-3/4; 9-13 years = grades 4/5-8; 14-19 years = grades 9-12).

[8] Studies 1-4 used a similar multiple regression approach were year-by-year standardized regression weights were plotted (as a function of age) and a smooth polynomial curve fit to the complete set of weights across all ages. For the current review we divided the smoothed curves into the three broad age groups and classified the IVs as significant if the smoothed population parameter curve estimates were significant *50% or more* of the time across the broad age group.

[9] The initial goal was to conduct a meta-analysis with a common effect size metric. Due to the limited number of research investigations, the lack of sufficient reported statistics for the calculation of effect sizes in some manuscripts, lack of independence of samples, and variety of statistical procedures used and results reported, the use of a common meta-analytic effect size metric was deemed impossible. Instead, a *significant relationship was defined as significant (t-test; F-test; regression weight; SEM effect size) if so reported by the studies researchers*. No attempt was made to separate *direct* and *indirect* (see Keith, 2006) effects for IVs—an IV was considered significant if the combined *total* (direct + indirect) effect was deemed significant in each respective analysis. In

analyses where only direct IV effects were reported, we calculated the indirect effects and then used the combined total effect value to determine if a significant relation was present. In analysis that reported split-sample model calibration and validation samples for the same age (e.g., Floyd et al., 2007), only the results from the cross-validated validation sample were used to determine significance. Keith (1997) reported separate IV effect sizes for white, black and Hispanic samples. Via multiple group confirmatory factor analysis (MGCFA), Keith (1997) reported the models to be invariant across groups. The median effect sizes across the three subsamples were used for significance classification in this review.

[10] *Tentative* or *speculative* results were those that were: (a) between 20-29% in consistency, (b) were based on a very small number of analyses (e.g., n=2), and/or (c) were based only on McGrew's (2007) exploratory multiple regression analysis of manifest WJ III variables at the individual IV test level.

[11] As indicated in Table 1 and all Figures, broad Gc was coded as consisting of the combination of the narrow abilities of language development (LD) and lexical knowledge (VL). This reflects the fact that the WJ-R Gc cluster and the WJ III Verbal Comprehension test, which were the primary Gc variables included in this review, have been noted to be combined measures of LD and VL (see Flanagan, Ortiz & Mascolo, 2001, 2007; also see McGrew & Woodcock, 2001).

[12] When discussing possible age-related or developmental trends, it is important to note that it is the *consistency of significance* that is being discussed.

[13] Page limitations did now allow for the presentation of selective referral-focused intelligence testing based on the findings of the current review. McGrew (2009) has presented examples of how to use the results of the current review to complete such assessments. A on-line PowerPoint slide show can be viewed at: *http://www.slideshare.net/iapsych/chc-selective-referralfocused-assessment-scenarios*