

## Against coefficient of variation for estimation of intraindividual variability with accuracy measures

Philippe Golay Delphine Fagot Thierry Lecerf  
University of Geneva

Previous studies have shown that intraindividual variability (*iV*) in performance is an important indicator of individual's cognitive functioning and neurological integrity. While most experiments have examined *iV* of performance using Reaction Time data (RTs), few studies have considered it with accuracy measures (e.g. number / percentage of correct responses). For these two types of measures, intraindividual standard deviation (*iSD*) or intraindividual coefficient of variation (*iCV*; intraindividual standard deviation divided by the individual mean) were used as indicators of *iV* in performance. However, because accuracy data have a lower and an upper bound (in contrast to RTs), we illustrate both formally and with simulated data, that the *iCV* cannot be used with accuracy measures. We also show that the coefficient *iCV* is influenced by the number of items which is an issue when dealing with missing data. We further provide formulas that may help researchers to visualize and correctly interpret their data using any spreadsheet software. The current article finally proposes an alternative coefficient ( $\zeta$ ) to examine *iV* in performance with accuracy measures that shows similar behaviour as does *iCV* with RTs data.

Studies in classical experimental psychology have mainly focused on level of performance. From this perspective, intraindividual variability of performance

---

P. Golay and T. Lecerf, Faculty of Psychology and Educational Sciences, University of Geneva, Switzerland and Distance Learning University, Switzerland; D. Fagot, Faculty of Psychology and Educational Sciences, University of Geneva, Switzerland and Center for interdisciplinary Gerontology, University of Geneva, Switzerland.

The authors thank their colleagues of the Developmental and Differential Psychology group of the University of Geneva for stimulating discussions. The authors also thank Denis Cousineau for his insightful comments on an earlier version of this manuscript.

Correspondence concerning this article should be addressed to Philippe Golay, FPSE -Psychology, University of Geneva, 40, Bd. du Pont d'Arve, CH-1205 Geneva, Switzerland, E-Mail: philippe.golay@unige.ch, Tel : +41 22 3799237

(hereafter intraindividual variability, *iV*) was considered as "noise" or as measurement error (de Ribaupierre, 1993; 1998; Salthouse & Nesselrode, 2010). However, during the last two decades, evidence has accumulated that *iV* might play an important role in the study and the comprehension of cognitive behavior and cognitive development. *iV* is a "signal" and not a "noise" (Nesselrode & Salthouse, 2004). It has been shown for instance that *iV* is related to level of ability for several cognitive domains (Jensen, 1992; Hultsch, MacDonald & Dixon, 2002). Nesselrode and Salthouse (2004, p. 49) wrote for instance "...intraindividual variability is a valid indicator of substantively important events". Thus, the question today is not whether *iV* should be examined as an indicator of cognitive behavior but how it could be studied. Indeed, several methodological questions remain to be addressed. For instance, Schmiedek, Lövdén and Lindenberger (2009) addressed the question of the assumption of the lack of relation between means and variances. Wagenmakers and Brown (2007) investigated the relation between mean RTs and standard deviation RTs and

reported that this relation was most often linear. Nesselroade and Salthouse (2004) on their part compared the magnitude of  $iV$  to the magnitude of between-person variability, while Boker, Molenaar and Nesselroade (2010) addressed the question of the time scale for measurement of  $iV$ . Another more general issue concerns the measurement: which scores might be used to index  $iV$  in performance (Allaire & Marsiske, 2005); this is the main goal of this paper.

Within the differential psychology approach, three types of variability can be examined. The first type of difference is focused on the differences between individuals. This type of difference is called “interindividual differences or *diversity*” (Hale, Myerson, Smith & Poon, 1988; Hultsh & MacDonald, 2004). Second, differential psychology examined intraindividual variability across tasks, called “*dispersion*” (Nesselroade, 1991; Shammi, Bosman & Stuss, 1998). In that case, multiple tasks are administered to a single individual. Finally, one can examine intraindividual variability across occasions or across trials. This type of intraindividual variability is labelled *inconsistency*. These last two types of variability concern the variability within individuals. It should be noted that inconsistency or fluctuations in individual performances within one task is the most often studied phenomenon. This type of intraindividual variability (*inconsistency*) is the main focus of this study and will be further referred as  $iV$ .

Interest in  $iV$  has increased over the past two decades, and more particularly within the cognitive aging domain. Several studies were conducted following this approach to describe, understand or even predict cognitive change with age (Hultsch & MacDonald, 2004; Lindenberger & von Oertzen, 2006; Lövdén, Li, Shing & Lindenberger, 2007; MacDonald, Hultsch & Dixon, 2003; Nesselroade, 1991). Many authors showed that  $iV$  increases with age (Bunce, MacDonald & Hultsch, 2004; Hogan, 2003; Hultsch, MacDonald & Dixon, 2002; Lindenberger & von Oertzen, 2006). In addition, authors reported increased  $iV$  in different pathologies affecting the central nervous system, including schizophrenia (Winterer & Weinberger, 2004), traumas (West, Murphy, Armilio, Craik, & Stuss, 2002), dementia (Hultsch, MacDonald, Hunter, Levy-Bencheton & Strauss, 2000) or Parkinson disease (Burton, Strauss, Hultsch, Moll & Hunter, 2006). These findings indicated that an increase of  $iV$  could reflect a decrease in the integrity of the central nervous system and can be considered as a risk factor for mortality (Stuss, Murphy, Binns & Alexander, 2003; West et al., 2002). It should be mentioned that higher degree of  $iV$  could be positive for some circumstances or for some variables; in other words, an elevated  $iV$  is not necessarily negative (Allaire & Marsiske, 2005). Evidence concerning the

saliency of  $iV$  was also provided by studies conducted with children. It was demonstrated for instance that children with ADHD showed higher  $iV$  when compared to typically developing children (Geurts et al., 2008; Leth-Steensen, Elbaz, & Douglas, 2000). Finally, other studies demonstrated that  $iV$  was more important with children compared to young adults (see Neuringer, 2002 for a review). In summary, whichever period of the lifespan, data consistently suggest that  $iV$  may be a “trait”, a “character” of individuals, which can show a relatively long-term stability in adulthood (Bielak et al., 2010). Thus,  $iV$  must be taken into account to improve understanding of the development and / or of the complexity of human behavior.

The magnitude of  $iV$  is classically evaluated by calculating the intraindividual standard deviation ( $iSD$ ) and / or the intraindividual coefficient of variation ( $iCV$ ;  $iSD$  divided by individual mean). However, both have statistical problems, and hence it has been suggested that these two scores provide an incomplete understanding of  $iV$  and are not completely adequate (Van Geert & Van Dijk, 2002). For instance, Deboeck, Montpetit, Bergeman and Boker (2009) demonstrated that standard deviation has two main limits: it neglects the ordering of the observations and the time scaling (day, week, month, etc). In our view, these scores have another limitation: no distinction was made between latency and accuracy measures. While most experiments were conducted on reaction time (RTs) tasks, few studies were conducted on accuracy data (e.g. number / percentage of correct responses). Nevertheless, it was implicitly assumed that whatever the nature of the data (RTs or accuracy) both indices could be used. Here we will show that  $iCV$  cannot be used with accuracy measures.

#### **Measures of intraindividual variability and their relationships with RTs or Accuracy data**

Multiple indices may be computed as measures of  $iV$  (Hultsch et al., 2000; Slifkin, & Newell, 1998). However, one common way to examine  $iV$  is by calculating the intraindividual standard deviation ( $iSD$ ). To examine inconsistency, this score is generally computed across trials for each individual. The result is expressed in the same metric as the measured data (i.e.  $iSD$  in ms if RTs are measured in ms). Furthermore, most of the studies have shown that  $iSD$  is very sensitive to the mean, and hence is strongly correlated with the mean RTs: lower RTs are associated with lower  $iSD$  and higher RTs are associated with higher  $iSD$ . This is generally considered to be a problem when comparing  $iV$  between groups with large differences in mean performance. Therefore, it is difficult to compare  $iSD$ s without taking the mean into account. In order to deal with the problem of the correlation between

mean and  $iSD$ , the intraindividual coefficient of variation is frequently used ( $iCV$ ; ratio of the  $iSD$  by the individual mean). The  $iCV$  could be considered as relative or intrinsic variability (Lewontin, 1966) and is a dimensionless number; thus, different samples could be compared.  $iCV$  is often multiplied by 100 and expressed as percentage which, we will see, can sometimes be misleading since the upper bound of this coefficient is rarely 1 or 100% but depends on the number of items.

In the following section, we will compare  $iSD$  and  $iCV$  for RTs first and then for accuracy measures. We will demonstrate that  $iSD$  and  $iCV$  have different profiles depending on the nature of the dependent variables (RTs /accuracy). Finally, in order to solve the problems raised by the use of  $iCV$  with accuracy scores, we suggest an adjusted coefficient ( $\zeta$ ) to estimate  $iV$  with such data. Thus, the main goal of this study was to illustrate to what extent  $iV$  scores behave differently with RTs or accuracy data. Keeping with this goal, we propose a way to enhance the interpretation of the data by proposing functions that can be plotted conjointly with the results of the experiment. We will also show that because task difficulty is a very important factor with accuracy tasks, there is a need to take into account this variable. In contrast, it is not necessary to take task difficulty into account with RTs measures.

#### Standard deviation from lower-bounded-only data (RTs)

We first examined characteristics of RTs distributions and the relationship between  $iV$  and intraindividual mean ( $iM$ ). With RTs tasks, data could theoretically vary from zero to any positive number. In other words, the lower bound of RTs is zero, but there is no real upper bound. Thus, RTs could be defined on a semi-finite interval. In that case,  $iSD$  reaches a maximum in a "Monopoly situation" (all cases have zero values save one; Dodd, 1952). The higher bound of  $iSD$  as a function of  $iM$  can be described by the following equations where  $n$  is the number of items:

$$\begin{aligned} \max(iSD) &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - iM)^2} \\ &= \sqrt{\frac{1}{n-1} \left( \sum_{i=1}^{n-1} (0 - iM)^2 + (x_n - iM)^2 \right)} \end{aligned}$$

We can further show that the mean ought to be the free datum ( $x_n$ ) divided by  $n$ :

$$\begin{aligned} iM \text{ is by definition } &\frac{1}{n} \left( \sum_{i=1}^{n-1} x_i \right) \\ &= \frac{1}{n} \left( \sum_{i=1}^{n-1} 0 + x_n \right) \\ &= x_n/n \end{aligned}$$

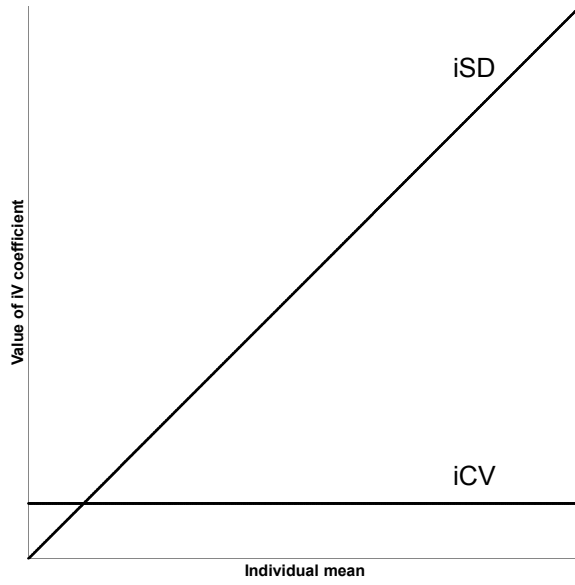
so that the free datum is  $n$  times the mean. Returning to  $\max(iSD)$ , we get:

$$\begin{aligned} \max(iSD) &= \sqrt{\frac{1}{n-1} ((n-1)iM^2 + (niM - iM)^2)} \\ &= \sqrt{\frac{1}{n-1} ((n-1)iM^2 + ((n-1)iM)^2)} \\ &= \sqrt{\frac{1}{n-1} ((n-1)iM^2 + (n-1)^2iM^2)} \\ &= \sqrt{\frac{1}{n-1} ((n-1) + (n-1)^2)iM^2} \\ &= \sqrt{\frac{1}{n-1} ((n-1)(1+n-1))iM} \\ &= \sqrt{n}iM \end{aligned}$$

This formula can also be written as  $\max(iSD) = iM\sqrt{n-1}$  when starting from the population-standard deviation formula. The maximum value of  $iSD$  is a function of both the number of items and the mean. Figure 1 shows that the maximum value of  $iSD$  with RTs tasks linearly increases as a function of  $iM$ . This situation implies a variance compression because there is not much room for variability when the value of the individual mean ( $iM$ ) is small.

#### Coefficient of variation from lower-bounded-only data (RTs)

Because of the positive correlations between  $iV$  and  $iM$ , authors have proposed to statistically control for RTs mean when examining  $iV$ . It is the reason why  $iCV$  was proposed. The higher bound of  $iCV$  can easily be computed from the first equation. By dividing  $iSD$  by  $iM$ , the equation simply becomes:  $\max(iCV) = \sqrt{n}$ . The upper bound of  $iCV$  is a constant which is not necessarily 1 or 100%. The values of  $iCV$  are therefore not a function of  $iM$  but are dependent of the number of items. This relationship suggests that missing data is an important issue: the variability could be underestimated but there would be no such effect on the value of  $iM$ . The comparison of  $iCV$ s obtained from experiments involving a different number of items (or sample size) should only be made with caution all the more so if the number of items is small (Martin & Gray, 1971).

Figure 1. Reaction time data: Maximum value of  $iSD$  and  $iCV$  as a function of individual mean

#### Standard deviation from lower and upper bounded data (Accuracy data)

Accuracy data on the other hand are measured by the total number of correct answers on any given trial. If participants can remember 3 out of 5 words from a list, they will get a score of 3 out of 5 (or 60%). Accuracy data are therefore defined on a finite interval with two predefined lower (0%) and upper bounds (100%). Let's call the minimum value  $x_1$  and the maximum value  $x_2$ .  $D$  is the score range which is computed as  $x_2 - x_1$ . With accuracy data, the maximum value of  $iSD$  is reached when the data is split to maximal bimodality (half of the data have the minimum value and the other half have the maximum value). In the extreme bimodality scenario we always have  $iM = (x_1 + x_2)/2$ . The distance between  $x_1$  or  $x_2$  to  $iM$  is  $D/2$ . The number of items is still named  $n$ .

Therefore we can write:

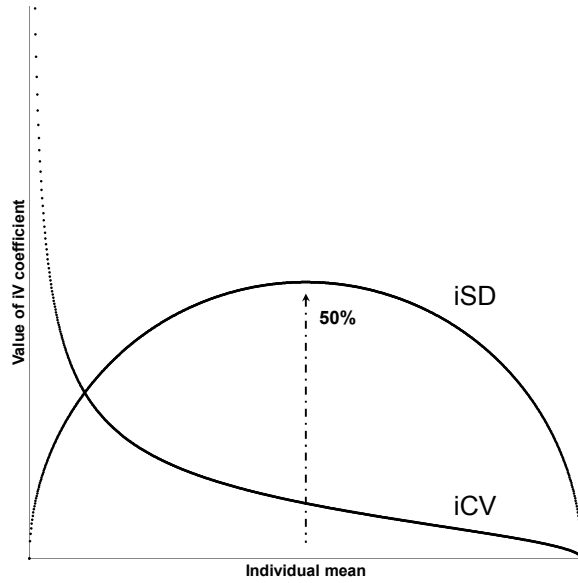
$$\begin{aligned} \max(iSD) &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_1 - iM)^2} \\ &= \sqrt{\frac{1}{n-1} \left( \sum_{i=1}^{n/2} (x_1 - iM)^2 + \sum_{i=n/2+1}^n (x_2 - iM)^2 \right)} \\ &= \sqrt{\frac{1}{n-1} \left( \frac{n}{2} (D/2)^2 + \frac{n}{2} (D/2)^2 \right)} \end{aligned}$$

$$= \sqrt{\frac{n}{n-1} D^2 / 4}$$

We assume that the score range  $D$  is always positive so we can further write  $\max(iSD) = \frac{D}{2} \sqrt{\frac{n}{n-1}}$ . In this scenario,  $\max(iSD)$  is a constant that only depends on the number of items and the score range  $D$ .

However we can further generalize the extreme bimodality scenario for other values than  $iM = (x_1 + x_2)/2$ . This generalized bimodality scenario is a situation in which all the data are still either at the minimum or at the maximum value, thus the standard deviation with respect to  $iM$  is locally maximized. However, the ratio between these two extreme values is allowed to vary: the number of items that needs to be equal to respectively  $x_1$  and  $x_2$  are variables and are calculated from  $iM$ . Accordingly, since the value of  $iM$  is no longer fixed to  $(x_1 + x_2)/2$  we can assume that the maximum value of  $iSD$  will be lower than the previously established upper limit  $(\frac{D}{2} \sqrt{\frac{n}{n-1}})$ .

The number of items that needs to be equal to respectively  $x_1$  and  $x_2$  for a given  $iM$  in order to get the maximum  $iSD$  is calculated as follows: number of minimum values  $nb(x_1) = n - n iM / (x_2 - x_1)$  and number of maximum values  $nb(x_2) = n iM / (x_2 - x_1)$ . It is easy to show that  $nb(x_1) + nb(x_2) = n$ . Getting back to  $\max(iSD)$  we can write:

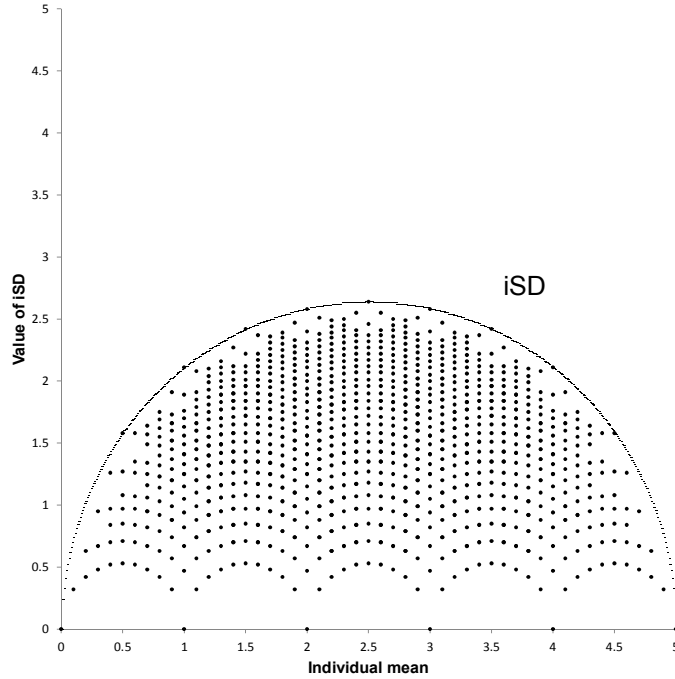
Figure 2. Accuracy data: Maximum value of  $iSD$  and  $iCV$  as a function of individual mean

$$\begin{aligned}
 \max(iSD) &= \sqrt{\frac{1}{n-1} \left( \sum_{i=1}^{n-\frac{n \cdot iM}{x_2-x_1}} (x_1 - iM)^2 + \sum_{i=n-\frac{n \cdot iM}{x_2-x_1}+1}^n (x_2 - iM)^2 \right)} \\
 &= \sqrt{\frac{1}{n-1} \left( \left( n - \frac{n \cdot iM}{x_2-x_1} \right) (x_1 - iM)^2 + \left( \frac{n \cdot iM}{x_2-x_1} \right) (x_2 - iM)^2 \right)} \\
 &= \sqrt{\frac{1}{n-1} n (-iM^2 + iM(x_2 - x_1) + x_1^2)} \\
 &= \sqrt{\frac{n}{n-1} (-iM^2 + iM(x_2 - x_1) + x_1^2)} \\
 &= \sqrt{\frac{n}{n-1}} \sqrt{-iM^2 + iM(x_2 - x_1) + x_1^2}
 \end{aligned}$$

Or more simply:  $\max(iSD) = \sqrt{\frac{n}{n-1}} \sqrt{iM D - iM^2}$ , when  $x_1 = 0$  and  $(x_2 - x_1)$  is referred as  $D$ . This formula can also be written as  $\max(iSD) = \sqrt{iM D - iM^2}$  when starting from the population-standard deviation formula. In that case the maximum value of  $iSD$  is a function of  $iM$ , score range ( $D$ ) and the number of items. Let us remind that  $D$  is defined as the maximum number of correct responses minus the minimum number of correct responses. Its value is defined as 1 when accuracy scores are expressed in percent (%) of correct answers. Figure 2 shows that the  $\max(iSD)$  function follows an arc-shaped curve reaching its absolute maximum value when  $iM$  is equal to 50% of correct responses. This absolute maximum value of  $iSD$  was previously established in the extreme bimodality scenario. It is the global maximum of the  $\max(iSD)$  function.

Because of the lower and upper bounds of accuracy scores, the maximum value of  $iSD$  will not monotonically increase as the value of  $iM$  gets bigger (as it does for RTs).  $\max(iSD)$  will in fact decrease when  $iM$  is greater than 50%. Maximum values of  $iSD$  with accuracy data are dependent on the individual mean. Most importantly, it will not be possible to observe much variability at all when the values of  $iM$  are either close to zero or close to upper bound (100%) (Figure 2). This can be referred as variance compression on both sides.

Participants with very low or very high scores do not have much room to show variability. Furthermore, if most participants get very high or very low scores,  $iM$  and  $iSD$  will be heavily correlated. In such situations,  $iM$  and  $iSD$  will carry almost the same information. It is crucial to adjust carefully the task difficulty to avoid this phenomenon. This problem is not readily visible on a XY-type scatter plot if the  $\max(iSD)$  curve is not drawn but will eventually get remarked through a very high degree of correlation between mean and variability scores. We argue that plotting  $\max(iSD)$  together with the experimental data can be a great help to determine the reason of dependency in the data. It is easily done with any spreadsheet software by plotting the  $\max(iSD)$  function on the desired range.

Figure 3.  $iSD$  and individual mean for 10 items scored between 0 and 5

#### ***Coefficient of variation from lower and upper bounded data (Accuracy data)***

Finally, the behaviour of  $iCV$  with accuracy scores can also be described as a function of  $iM$ , score range ( $D$  which is equal to  $x_2 - x_1$ ) and the number of items  $n$ :

$$\begin{aligned} \max(iCV) &= \frac{\max(iSD)}{iM} \\ &= \frac{\sqrt{\frac{n}{n-1}} \sqrt{(-iM^2 + iM(x_2 - x_1) + x_1^2)}}{iM} \\ &= \frac{\sqrt{\frac{n}{n-1}} \sqrt{iMD - iM^2}}{iM} \end{aligned}$$

The most notable feature on Figure 2 is that the values of  $iCV$  will monotonically decrease as  $iM$  increases. The calculation of  $iCV$  on accuracy data will yield values that are greatly dependent on the values of  $iM$ . Let us remind that this was not the case with RTs data where the upper bound of  $iCV$  was constant and showed no direct relationship with  $iM$  (see Figure 1). Therefore we consider that  $iCV$  is not an adequate  $iV$  coefficient with accuracy scores.

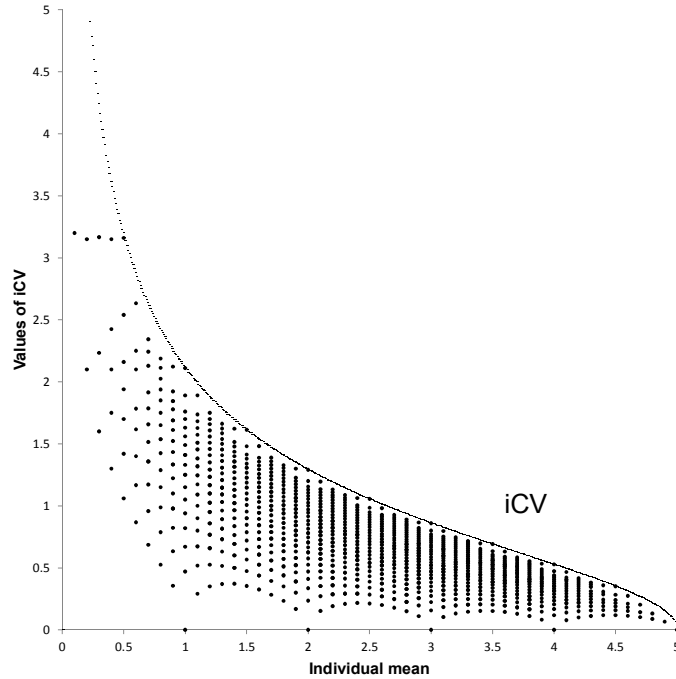
Previous equations are simple algebraic maximization of the value of  $iSD$  with respect to the value of  $iM$  on semi-finite and finite intervals. They describe the behaviour of both coefficients in a very general formal manner and can easily be drawn when plotting mean performance and variability data. We next illustrate the previous findings

with data that are much more discrete in nature (i.e. in an experiment involving a small number of items and a small score range). Figure 3 shows every possible value of  $iSD$  in relation to  $iM$  with simulated data involving 10 items (each scored from 0 to 5). All possible results have been generated. Due to the discrete nature of accuracy data, the  $6^{10}$  possible results converged to about 500 individual dots on the graphic. One should notice that due to the small number of items and the discrete nature of most accuracy scores, the theoretical maximum values of  $iSD$  are not always reached.

Figure 4 presents the exact same dataset with  $iCV$  instead of  $iSD$ . Again, all dots lie into the area described by the  $\max(iCV)$  equation. Note that the maximum values are also rarely reached. With these simulated data, because  $iCV$  necessarily decreases when the value of  $iM$  increases, both are heavily correlated and share here about 57% of their variance. Note that the shape of the scatter and the correlation between the performance and variability coefficients are only dependent on the numerical boundaries of  $iCV$ . The linear dependency between  $iCV$  and  $iM$  could have easily been misinterpreted if the  $\max(iCV)$  curve had not been drawn.

#### ***Adjusted coefficient of variation from lower and upper bounded data (Accuracy data)***

Finally, it is possible to create a coefficient  $\zeta$  that will show the same behaviour as does  $iCV$  on a semi-finite

Figure 4. *iCV* and individual mean for 10 items scored between 0 and 5

interval (i.e.: a constant, flat upper bound that is not related to  $iM$ ).  $\zeta$  can be defined as the ratio between variability ( $iSD$ ) and the maximum variability that is possible to reach at that given level of performance ( $\max(iSD)$ ). The range of this coefficient is between 0 to 1 and  $\zeta$  is a dimensionless number.

$$\begin{aligned}\zeta &= \frac{iSD}{\max(iS)} \\ &= \frac{iSD}{\sqrt{\frac{n}{n-1} \sqrt{(-iM^2 + iM(x_2 - x_1) + x_1^2)}}} \\ &= \frac{iSD}{\sqrt{\frac{n}{n-1} \sqrt{iM D - iM^2}}}\end{aligned}$$

As described on Figure 5, the values of  $\zeta$  show no linear relationship with  $iM$ . This behaviour is similar to  $iCV$  with RTs data (lower bounded-only data), in which there is also no relationship between the variability coefficient and the mean. The resolution close to the lower and upper bound is nevertheless scarce and suggests it is not possible to get reliable information about variability at the extremes. Therefore, task difficulty should still be carefully adjusted.

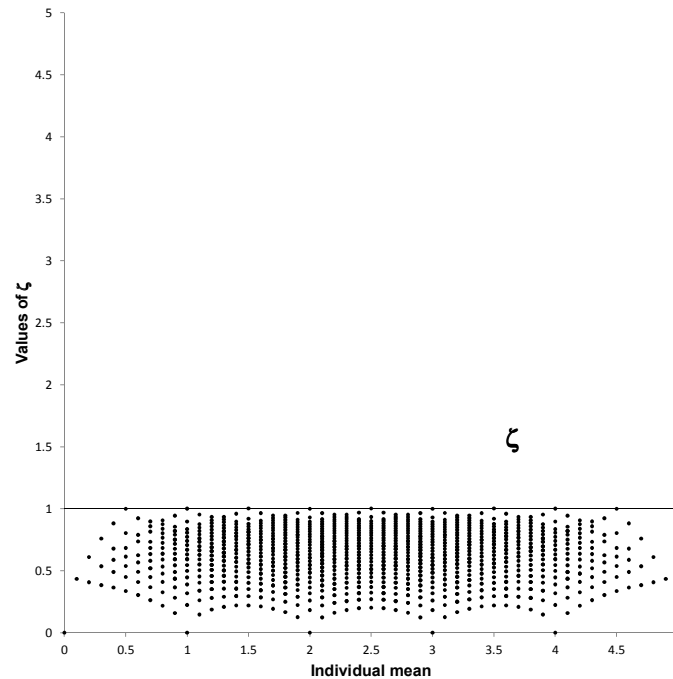
### Discussion

The goal of this study was to illustrate that index scores of  $iV$  with accuracy data and RTs are not *de facto* equivalent. RTs are defined on semi-finite intervals with lower but no upper bounds. On the opposite, Accuracy data are defined between a minimum and a maximum score.  $iV$  on accuracy

data is sensitive to the task's difficulty level because the maximum value of  $iSD$  is constrained when the ratio of correct response is either very high or very low. Both coefficients could become heavily correlated in such situations. It has been well known for quite a long time that one should adjust the task for difficulty to avoid ceiling and floor effects when studying mean level of performance. Most importantly, in the present study we demonstrated that it is also crucial to carefully adjust the task difficulty when the focus is put on performance variability with accuracy measures.

The  $iCV$  does not show a similar behaviour when applied on RTs and accuracy data. With RTs, the  $iCV$  provides a measure of relative variability with a constant upper bound that is related to the total number of items involved. With accuracy data, the upper bound of  $iCV$  will inevitably tend to zero as  $iM$  increases. This is problematic for several reasons. First, values of  $iCV$  will behave in a totally different manner with RTs and accuracy scores and researchers should be well aware of this fact. Second,  $iCV$  will almost always depict participants with lower accuracy scores as being more variable as well:  $iCV$  will share a very large amount of variance with the mean. Thus it could be a very redundant and potentially misleading indicator of  $iV$ .

Finally, it is possible to build a coefficient  $\zeta$  with the same upper-bound characteristics as  $iCV$  with RTs data.  $\zeta$  can be written as the ratio between variability ( $iSD$ ) and the maximum variability ( $\max(iSD)$ ) attainable at that given

Figure 5.  $\zeta$  and individual mean for 10 items scored between 0 and 5

level of performance. The latter relevance of this coefficient to the study of *iV* with accuracy measures should nevertheless be further studied.

### References

- Allaire, J., & Marsiske, M. (2005). Intraindividual Variability May Not Always Indicate Vulnerability in Elders' Cognitive Performance. *Psychology and aging*, 20(3), 390.
- Bielak, A. A. M., Hultsch, D. F., Strauss, E., MacDonald, S. W. S., & Hunter, M. A. (2010). Intraindividual variability in reaction time predicts cognitive outcomes 5 years later. *Neuropsychology*, 24(6), 731.
- Boker, S. M., Molenaar, P., & Nesselrode, J. R. (2009). Issues in intraindividual variability: Individual differences in equilibria and dynamics over multiple time scales. *Psychology and aging*, 24(4), 858.
- Bunce, D., MacDonald, S. W. S., & Hultsch, D. F. (2004). Inconsistency in serial choice decision and motor reaction times dissociate in younger and older adults. *Brain and Cognition*, 56(3), 320-327.
- Burton, C. L., Strauss, E., Hultsch, D. F., Moll, A., & Hunter, M. A. (2006). Intraindividual Variability as a Marker of Neurological Dysfunction: A Comparison of Alzheimer's Disease and Parkinson's Disease. *Journal of Clinical and Experimental Neuropsychology*, 28(1), 67-83.
- de Ribaupierre, A. (1993). Structural invariants and individual differences: On the difficulty of dissociating developmental and differential processes. In R. Case & W. Edelman (Eds.), *The new structuralism in cognitive development: Theory and research on individual pathways. Contributions to human development* (Vol. 23, pp. 11-32). Basel, Switzerland: Karger.
- de Ribaupierre, A. (1998). Développement cognitif et différences individuelles. In M. Sabourin, F. I. M. Craik & M. Robert (Eds.), *Advances in Psychological Sciences* (Vol. 2: Biological and cognitive aspects, pp. 531-555). Hove, UK: Psychology Press Ltd.
- Deboeck, P. R., Montpetit, M. A., Bergeman, C., & Boker, S. M. (2009). Using derivative estimates to describe intraindividual variability at multiple time scales. *Psychological methods*, 14(4), 367.
- Dodd, S. (1952). On percentage moments: Interpreting the moments as percents of their maximums. *The Scientific Monthly*, 74, 220-222.
- Geurts, H. M., Grasman, R. P. P., Verté, S., Oosterlaan, J., Roeyers, H., van Kammen, S.M., et al. (2008). Intraindividual variability in ADHD, autism spectrum disorders and Tourette's syndrome. *Neuropsychologia*, 46, 3030-3041.
- Hale, S., Myerson, J., Smith, G. A., & Poon, L. W. (1988). Age, variability, and speed: between-subjects diversity. *Psychology and Aging*, 3(4), 407-410.
- Hogan, M. J. (2003). Average speed or variable speed--what do older adults really need? *Irish Journal of Psychology*, 24(3-4), 161-183.
- Hultsch, D. F., & MacDonald, S. W. S. (2004). Intraindividual



- variability in performance as a theoretical window onto cognitive aging. In R. A. Dixon, L. Bäckman & L.-G. Nilsson (Eds.), *New frontiers in cognitive aging* (pp. 65–88). Oxford, UK: Oxford University Press.
- Hultsch, D. F., MacDonald, S. W. S., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 57B(2), P101-P115.
- Hultsch, D. F., MacDonald, S. W. S., Hunter, M. A., Levy-Bencheton, J., & Strauss, E. (2000). Intraindividual variability in cognitive performance in older adults: Comparison of adults with mild dementia, adults with arthritis, and healthy adults. *Neuropsychology*, 14(4), 588-598.
- Jensen, A. R. (1992). The importance of intraindividual variation in reaction time. *Personality and Individual Differences*, 13(8), 869-881.
- Leth-Steensen, C., Elbaz, Z. K., & Douglas, V. I. (2000). Mean response times, variability, and skew in the responding of ADHD children: A response time distributional approach. *Acta Psychologica*, 104, 167-190.
- Lindenberger, U., & von Oertzen, T. (2006). Variability in Cognitive Aging: From Taxonomy to Theory. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 297-314). New York, NY, US: Oxford University Press.
- Lewontin, R. (1966). On the Measurement of Relative Variability. *Systematic Zoology*, 15(2), 141-142.
- Lövdén, M., Li, S.-C., Shing, Y. L., & Lindenberger, U. (2007). Within-person trial-to-trial variability precedes and predicts cognitive decline in old and very old age: Longitudinal data from the Berlin Aging Study. *Neuropsychologia*, 45(12), 2827-2838.
- MacDonald, S. W. S., Hultsch, D. F., & Dixon, R. A. (2003). Performance variability is related to change in cognition: Evidence from the Victoria Longitudinal Study. *Psychology and Aging*, 18(3), 510-523.
- Martin, J., & Gray, L. (1971). Measurement of relative variation: Sociological examples. *American Sociological Review*, 496-502.
- Nesselroade, J. R. (1991). The warp and the woof of the developmental fabric. In R. Downs, L. Liben & D. S. Palermo (Eds.), *Visions of aesthetics, the environment, & development: The legacy of Joachim F. Wohlwill* (pp. 213-240). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Nesselroade, J. R., & Salthouse, T. A. (2004). Methodological and theoretical implications of intraindividual variability in perceptual-motor performance. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 59(2), P49
- Neuringer, A. (2002). Operant variability: Evidence, functions, and theory. *Psychonomic Bulletin & Review*, 9(4), 672-705.
- Salthouse, T. A., & Nesselroade, J. R. (2010). Dealing with short-term fluctuation in longitudinal research. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 65(6), 698.
- Schmiedek, F., Lövdén, M., & Lindenberger, U. (2009). On the relation of mean reaction time and intraindividual reaction time variability. *Psychology and aging*, 24(4), 841
- Shammi, P., Bosman, E., & Stuss, D. T. (1998). Aging and Variability in Performance. *Aging, Neuropsychology, and Cognition*, 5(1), 1 - 13.
- Slifkin, A. B., & Newell, K. M. (1998). Is variability in human performance a reflection of system noise? *Current directions in psychological science*, 7(6), 170-177.
- Stuss, D. T., Murphy, K. J., Binns, M. A., & Alexander, M. P. (2003). Staying on the job: The frontal lobes control individual performance variability. *Brain*, 126(11), 2363-2380.
- Van Geert, P., & Van Dijk, M. (2002). Focus on variability: New tools to study intra-individual variability in developmental data. *Infant Behavior and Development*, 25(4), 340-374.
- Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, 114, 830-841.
- West, R. L., Murphy, K. J., Armilio, M. L., Craik, F. I. M., & Stuss, D. T. (2002). Lapses of intention and performance variability reveal age-related increases in fluctuations of executive control. *Brain and Cognition*, 49(3), 402-419.
- Winterer, G., & Weinberger, D. R. (2004). Genes, dopamine and cortical signal-to-noise ratio in schizophrenia. *Trends in Neurosciences*, 27(11), 683-690.

Manuscript received 14 February 2012

Manuscript accepted 1 April 2012.