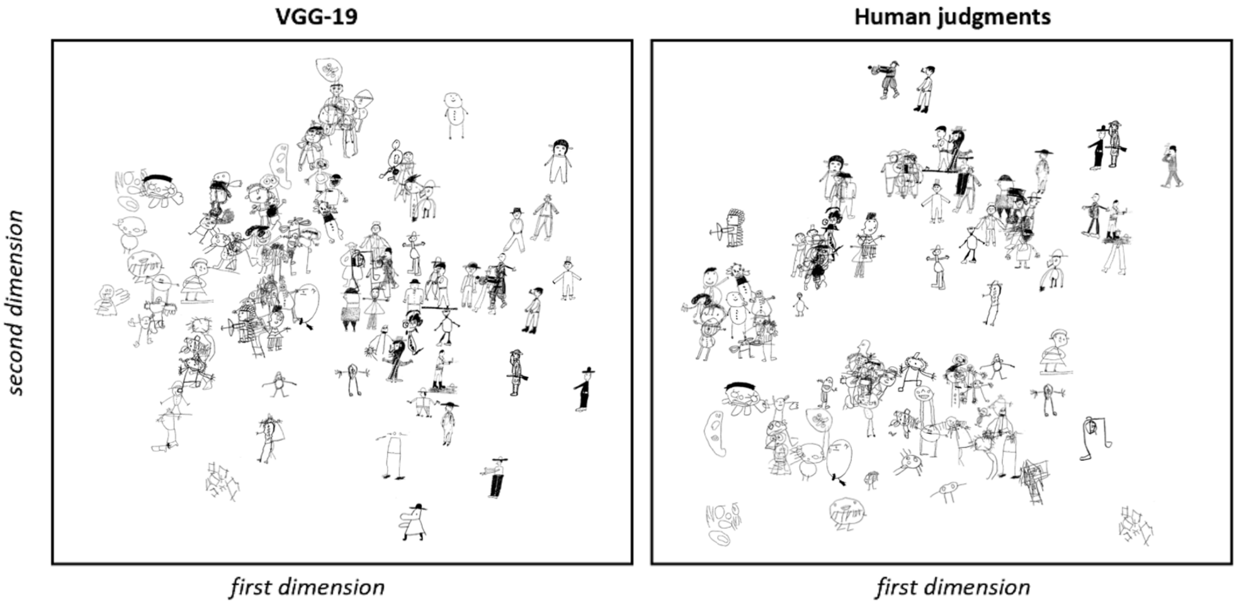
More than Goodenough: Finding latent structure in children's drawings through human perception and machine vision

Clint A. Jensen* (cjensen5@wisc.edu) University of Wisconsin-Madison
Timothy T. Rogers (ttrogers@wisc.edu) University of Wisconsin-Madison
Karl S. Rosengren (karl.rosengren@rochester.edu) University of Rochester

What drawings may reveal about a child's cognitive, emotional and physical development has a long history of interest, exploration, and conjecture. One of the most influential methodologies of characterizing children's drawings was outlined by Florence Goodenough, in her 1926 book, *Measurement of intelligence by drawings*. In her book, Goodenough outlines the Draw-A-Man test (DAMT), a checklist system to identify important features of human figures (e.g., body parts, facial features). The book provides both evidence of the DAMT as a non-verbal measure of intelligence and includes examples of human figure drawings that have been scored using the DAMT checklist of 46 standard features, with 5 additional items for human figure drawings depicted in profile. DAMT scores for a human figure drawing constitute the summed number of features present within an image. The resulting scores are then used along with a child's chronological age to determine a child's intelligence quotient (IQ), a measure that Goodenough believed was stable across one's life. While the use of drawings as a practical measure is not without controversy (Imuta et al., 2013), a recent large longitudinal study using the reduced 12-feature checklist protocol (Draw-A-Child; McCarthy, 1972) demonstrated the ability to significantly predict non-verbal intelligence from drawings produced at age 4 for the same participants at age 14 (Arden et al., 2014).

The present study builds upon recent findings that incorporate contemporary data science tools, including deep neural network models of vision and crowd-based similarity ratings to leverage latent structure present within children's drawings to predict demographic, motor, and other characteristics of the artist above and beyond the Draw-A-Child checklist (Jensen et al., 2022). Using the human figure drawings provided as training and practice examples within Goodenough's (1926) text as input, we describe findings using machine vision and/or human perception that embed those human figure drawings into low-dimensional space (See Figure 1). We demonstrate that such computational approaches provide additional explanatory variance to both the demographic measure of participant age, but also more critically to Goodenough's DAMT IQ scores. This outcome suggests that even in a small and highly controlled set of images, contemporary tools may represent a useful method to enhance drawing assessments (See Table 1). Further implications will be considered.

Figure 1. *Two-dimensional Embeddings*



Note. Two-dimensional embeddings for Goodenough (1926) human figure drawings based on VGG-19 vectors (left) vs human judgments of similarity (right). Each technique captures some aspects of the collection's structure. For VGG-19, circular shapes composed of light strokes are grouped in the left, while images that are comprised of darker/denser strokes appear toward the right. For human judgments figures at the bottom transition to more circular drawings and then to fuller depictions of the whole figure toward the top.

Table 1. *Comparison of Model Fits for Stepwise Regression*

Dependent Variable	*metric*	*n*	Goodenough DAMT Score		Draw-A-Child Score	Human 2D Embedding	VGG-19 2D Embedding
Age	adj. r^2	98	**0.58*****	**>**	**0.38*****	0.61	**0.64***
IQ	adj. r^2	98	**0.88*****	**>**	**0.50*****	**0.95*****	**0.94*****
	comparison model:		null		null	Goodenough DAMT Score	Goodenough DAMT Score

Note. Age models include gender and its interactions with other variables as regressors of no interest. The Goodenough Score model includes Goodenough's original checklist and the interactions of Age and Gender. Significance tests for these are against the null hypothesis, while the comparison signs (greater/less than) indicate whether one metric accounts for reliably more/less variance than another. Asterisks indicate significance levels at '*' $p < 0.05$ ', '**' $p < 0.01$, '***' $p < 0.001$.