The use of images in rating scales to assess attitudes, feelings and dispositions

Christine Merrell & Peter Tymms

17th Annual Meeting of the
Association for Educational Assessment - Europe
3rd – 5th November 2016

Contact: Christine.Merrell@cem.dur.ac.uk

Acknowledgement: This research was supported by a grant from the Department for Education, England.

Abstract

Across the social sciences, there are many latent variables which are of interest to researchers and others, including attitudes, social and emotional skills. Attempts to measure these commonly involve written questions and rating scales that require Likert-type responses. There are difficulties associated with using these types of measures with young children. Many young children have limited reading skills and vocabulary levels. They may also have limited conceptual understanding of the variables being measured. These difficulties can result in data which may not be reliable and which are not fairly comparable between children; Soto et al. (2011) have suggested that self-report questionnaires are not considered reliable for children under the age of 10 years. Historically, to get around these difficulties, adults (parents, teachers etc.) are often asked to complete assessments on the basis of their observations of children. However, these may not accurately capture the child's own feelings; children may not choose to show their feelings through overt behaviours and the adults may not have had the opportunity to amass sufficient information. Furthermore, if assessments of social and emotional skills are to be used in cross-cultural and international studies with children, translations are problematic with subtle nuances in wording at risk of being misinterpreted. A new approach to assessing these variables is needed.

One way forward is to make use of images and to ask children to pick out someone "who is like you" from a picture. The images can be carefully constructed to probe the construct of interest. This format of assessment can incorporate animations to more effectively convey a facial expression, body posture or a behaviour.

This paper describes the early development of a measure which uses images to assess the Big Five personality factors (openness, conscientiousness, extraversion, agreeableness, neuroticism). These have been identified by the OECD (2015) as being amongst the range of personal and social skills which are associated with positive outcomes in life and important to measure. It is vital to have a method to assess these behaviours which can be reliably used with young children from different cultures and backgrounds if we are to investigate differences between individuals and groups, and changes in response to interventions.

We share an idea of using animated images, which has shown promise for use with children under the age of 10 years, for whom existing methods of assessment are problematic. Ideas for further development, which build upon this initial thinking, are discussed.

Introduction

There is increasing interest in skills and attributes, over and above cognitive ability, which are changeable, and which are influential in shaping the lives of individuals. These include personal and social skills such as grit, self-control, curiosity, open-mindedness, motivation (Duckworth and Yeager, 2015). The term 'non-cognitive skills' is widely used as a descriptive title for such skills but it is considered to be problematic by many because these skills do involve some cognitive processing. Indeed, Carneiro et al., (2007) reported a positive association between cognitive and non-cognitive skills. However, as John Easton (2013) said: "*Everybody hates this term but everyone knows roughly what you mean when you use it and no one has a much better alternative*." Kautz et al. (2014) described non-cognitive skills as "*personal attributes not thought to be measured by IQ tests or achievement tests*". They go on to refer to the literature which suggests that these skills can be changed and develop in response to intervention.

The Big Five

Focusing in on the classification of non-cognitive skills, models have emerged and been refined over decades, and currently the Big Five model is widely regarded as an overarching framework. The Big Five dimensions are openness, conscientiousness, extraversion, agreeableness and neuroticism. Within these broad areas fit more narrowly defined skills such as grit, cooperation, competitiveness (Costa and McCrae, 1992; John and Srivastava, 1999; Kautz et al., 2014).

Costa and McCrae (1992) ascribed the following descriptors to the five dimensions:
Openness: Openness to experience, including 'fantasy, aesthetics, feelings, actions, ideas and values'.
Conscientiousness: 'Competence, order, dutifulness, achievement, striving, self-discipline, deliberation'.
Extraversion: 'Warmth, gregariousness, assertiveness, activity, excitement-seeking, positive emotion'.
Agreeableness: 'Trust, straightforwardness, altruism, compliance, modesty, tendermindedness'.
Neuroticism: 'Anxiety, hostility, depression, self-consciousness, impulsiveness, vulnerability to stress'.

Whilst this classification is considered to be useful and the words used to describe the dimensions are quite well understood, it is not without issues. John and Srivastava (1999) argued that we have yet to fully understand and explain the underlying processes and structures which manifest themselves in the behaviours for which we have applied descriptive labels.

Links to later outcomes

In their review of the predictive power of non-cognitive skills, Kautz et al. (2014) found them to explain a high proportion of the variance in achievement scores after taking account of IQ. However, they suggested caution when interpreting this, noting that cognitive development is not entirely independent of non-cognitive skills.

In addition to educationalists and psychologists, economists are interested in factors that predict later outcomes, in their case economic success. For example, Borghans et al. (2008) studied the relationship between personality factors, which they defined as "patterns of thoughts, feelings and behaviour" rather than the factors such as motivation, values, interests and attitudes which can influence personality. They found that personality factors are predictive of socioeconomic success whilst at the same time calling for more research into improved measures and acknowledging that there is more to be discovered about personality and other factors such as

motivation, and their relationship with later outcomes. Policy-makers around the world are engaging with the area; in 2014, education ministers and vice-ministers from several countries met at the OECD's informal ministerial meeting in Sao Paulo, Brazil, to discuss the question of which skills drive wellbeing and social progress (OECD, 2015). The OECD (2015) published a broad-ranging review of the importance of social and emotional skills for societal progress.

Issues with assessment

Whilst it is acknowledged that some non-cognitive skills can change in response to intervention, there are issues with the assessments used to measure these skills. Without good assessments, it is impossible to measure any changes and the impact of interventions. The issues are particularly salient for the assessment of young children. As noted above, Soto et al. (2011) have suggested that the results from self-report questionnaires can be considered as reliable for children aged 10 years and older but what about younger children?

Assessments of non-cognitive skills typically involve rating a level of agreement with descriptors on a Likert-type scale. Whilst these may be appropriate for adults, Merrell and Tymms (2016) described a range of issues associated with assessing young children. These included limited reading ability, concentration span and short-term memory. Also, it cannot be assumed that young children will have the conceptual understanding of the dimensions to be able to evaluate their own feelings and behaviours in an objective way. Their vocabulary acquisition may not be sufficiently sophisticated to enable them to both understand what is being asked and to respond appropriately. If we rely on adults to assess children on the basis of observations, there are further issues. If the adult is the child's parent, they will know their child well but may have limited experience of other children from across the population or they may bring their own bias into their ratings. A class teacher will be more likely than parents to have seen a wide range of children's behavior and may be more objective in their ratings but they may not know every child sufficiently well to be able to make reliable judgements.

Additional challenges are faced if an assessment is used across cultures and countries. Words can be interpreted in different ways and they can lose their nuance through translation. Paunonen et al. (1992) developed an assessment of personality using simple line drawings along with a five-point scale for the respondent to rate their agreement with how likely they were to behave like the character in the picture and this was compared with other established personality questionnaires (Paunonen, 2003). The scores from this nonverbal measure corresponded well with those from well-established questionnaires when used with a sample of college psychology students. Paunonen et al. (1992) trialled the nonverbal personality test alongside a verbal one with samples of adults in Canada, Finland, Poland and Germany, and found them to work well when used with samples of university graduates.

The aim of this study

We do want to be able to assess non-cognitive skills within young children if we are to monitor changes over time and in response to intervention within individuals, across groups and across cultures and countries. One potential way forward is to use figures within a picture scene and ask the child to pick out someone "who is like you". The scenes and the figures within them can be carefully constructed. Animated pictures could enhance understanding. This paper describes the early development of an assessment of the Big Five dimensions for use with young children. By using illustrations accompanied by very simple verbal instructions, this format of assessment aims to overcome some of the issues identified earlier with respect to children's reading ability etc.

This builds upon previous research by using more sophisticated illustrations and animations, and it has not, to our knowledge, been used before in an assessment of non-cognitive skills intended for young children.

Development of the items

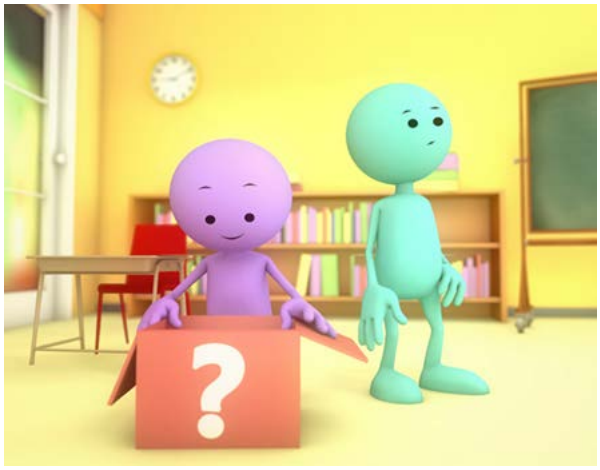
We created eighteen items which were intended to represent aspects of the Big Five dimensions, and an on-line assessment was built to trial those items. Our aim was to investigate the feasibility of this approach and we therefore limited the number of items developed and spanned the five dimensions rather than focusing on a single dimension in depth. This meant that we could consider the issues of trying to represent written descriptions of a range of dimensions through images. One item from each dimension is described below along with the corresponding screen-shots of the image.

The figures that we developed were intended to be gender and culturally neutral.

Each scene was accompanied by a verbal instruction to “pick the figure that is most like you”.

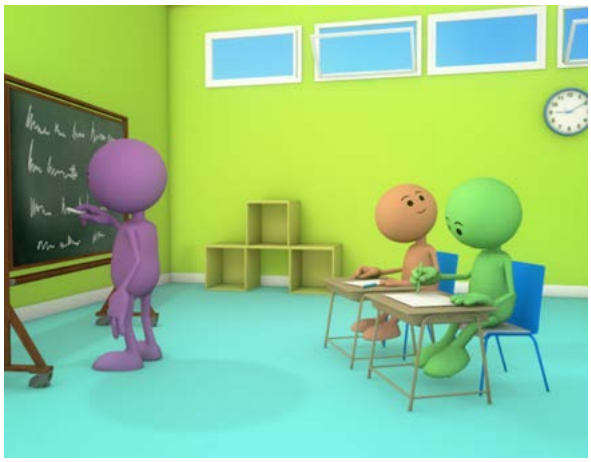
Openness

Here we try to represent openness to a new experience. One figure is looking inside a box with a pleased facial expression. The box has a question mark on the outside, which was intended to convey mysterious contents; a surprise to be explored. Another figure is looking disinterested and facing a different direction.



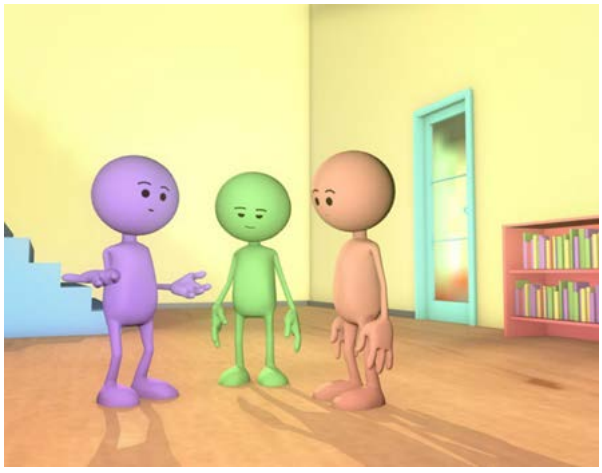
Conscientiousness

Here we have an animated scene that tries to represent dutifulness and self-discipline. Two figures are sitting at desks and a teacher is writing on the board. A fly buzzes into the classroom. One of the figures is more interested in watching the fly than concentrating on their class work.



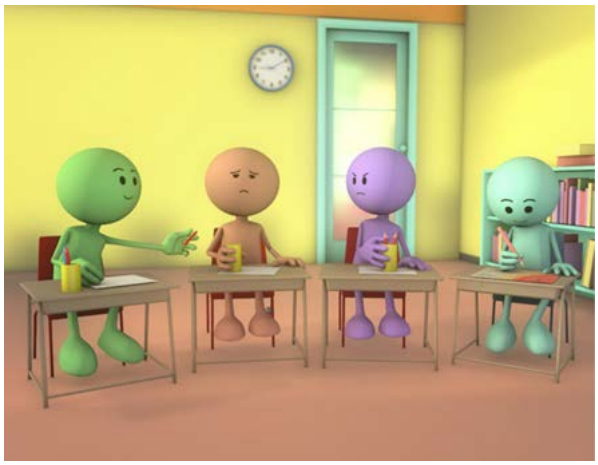
Extraversion

Here we have an animated scene of three figures in conversation where we try to convey gregariousness. One of the figures is leading the conversation, another is engaged periodically and a third appears not to want to participate.



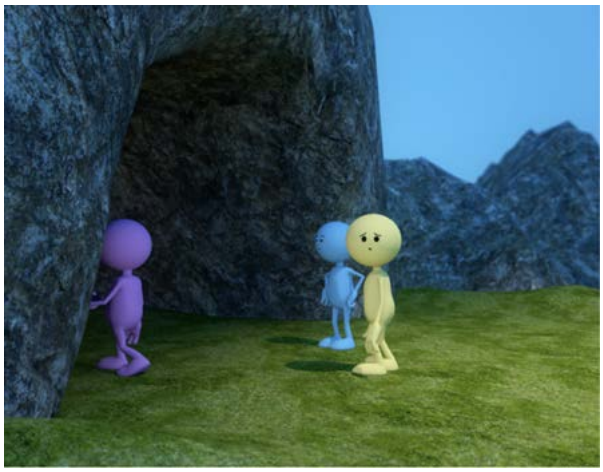
Agreeableness

In this animation, we try to convey altruism and tendermindedness. One figure tips up the pencil container and finds that it is empty. One of the neighbouring figures moves their pencil container away and clearly does not want to share. The figure at the other side offers a pencil to their neighbour. The figure at the end of the row is getting on with their work.



Neuroticism

In this animation, we try to convey anxiety. One figure shines their torch and enters the cave without hesitation. One figure hesitates but looks interested in going into the cave. The third figure steps back from the entrance and looks anxious.



See the Appendix for the full set of items.

Piloting the items

The project received ethics approval from the School of Education Committee at Durham University, and the items were piloted with a total of 48 participants (31 Females and 17 males) during March 2015. This was an opportunistic sample, intended to include a wide age range and amongst the young children to include a range of abilities but note that it is small and in no way intended to be representative of any particular groups at this stage of the assessment development. The age range included children aged 5 years through to adults. Twenty children were younger than age 10 at the time of the assessment. We were interested in looking at the responses of

children under the age of 10, where there is a gap in reliable and valid assessments of non-cognitive skills. We were also interested in comparing the responses from the younger children to those of older children and adults.

The assessment was administered on computer and took between five and ten minutes to complete.

The items were scored on a 1-3 scale such that the higher the ascribed value, the more the personality dimension had been endorsed. When a scene contained three figures, the one which was most closely aligned to the intended description scored 3 marks, the middle figure scored 2 marks and the figure least like the description scored 1 mark. When scenes contained only two figures, the figure which was most closely aligned to the description scored 3 and the one which was least closely aligned scored 1.

Analysis

Table 1 shows the percentage of responses within each item for the full sample, the children under 10, and children older than ten and adults.

Table 1

	Full Sample (n=48)			Children under age 10 (n=20)			Age 10 and older (n=28)		
Score	1	2	3	1	2	3	1	2	3
Openness 1	6%	N/A	94%	15%	N/A	85%	0%	N/A	100%
Openness 2	15%	N/A	85%	5%	N/A	95%	21%	N/A	79%
Openness 3	33%	N/A	67%	35%	N/A	65%	32%	N/A	68%
Openness 4	40%	N/A	60%	30%	N/A	70%	46%	N/A	54%
Extraversion 1	15%	54%	31%	5%	70%	25%	21%	43%	36%
Extraversion 2	11%	54%	35%	10%	30%	60%	11%	71%	18%
Agreeableness 1	17%	N/A	83%	18%	N/A	82%	16%	N/A	84%
Agreeableness 2	7%	29%	64%	0%	37%	63%	12%	23%	65%
Agreeableness 3	20%	2%	78%	18%	6%	76%	21%	N/A	79%
Conscientious 1	29%	N/A	71%	30%	N/A	70%	29%	N/A	71%
Conscientious 2	19%	42%	39%	15%	40%	45%	21%	43%	36%
Conscientious 3	23%	N/A	77%	30%	N/A	70%	18%	N/A	82%
Conscientious 4	40%	N/A	60%	40%	N/A	60%	39%	N/A	61%
Neuroticism 1	50%	44%	6%	75%	20%	5%	32%	61%	7%
Neuroticism 2	60%	27%	13%	75%	10%	15%	50%	39%	11%
Neuroticism 3	67%	23%	10%	70%	10%	20%	64%	32%	4%
Neuroticism 4	77%	10%	13%	60%	20%	20%	89%	4%	7%
Neuroticism 5	35%	44%	21%	55%	25%	20%	21%	57%	22%

Correlations

The correlations between items were investigated. They were predominantly positive but lower than 0.4. The highest correlations were between the items intended to assess neuroticism. This emphasized the difficulty of knowing, a priori, which aspect of personality would be picked up by items.

Scales

Rasch models were run for the five dimensions individually as well as for all of the items together. Not surprisingly, given the low correlations between items and the small number of items in each trait, the scales were of low reliability although the conscientiousness scale showed promise with a person reliability of 0.4, suggesting that if more items were available a reasonable scale could be created. The data imply that 17 similar items would be needed to produce a reliability of 0.7; with a refinement of the content of the items, fewer would be needed.

There was some evidence of differential item function by sex and age.

Factor analysis

There was insufficient data to produce decisive results in a factor analysis but a rotated solution, constrained to produce 5 factors resulted in a main component (16% of variance) which was mainly made up from the Openness and Neuroticism (negative) items.

Feedback from participants

At this stage in the development process, it is useful to interview the participants after they have completed the assessment. We focused on the younger children in the sample.

Some children under age 10 commented that they looked for the colour of the figure in the pictures and tried to choose a consistent colour of figure throughout rather than choosing a figure on the basis of their actions and facial expressions. Sometimes children chose the figure which corresponded to favourite colour. They thought that the different coloured figures had consistent traits, for example the yellow figure often appeared to be confident. This could be improved by using figures of the same colour but perhaps different depth of shade in addition to their different facial expressions and actions to emphasise the differences.

A small number of children age 6 commented that they liked to choose the figure whom they perceived to be 'doing the right thing'.

Individual comments were made about some of the scenes, for example one child, age 6, said they didn't understand the group of figures talking (Extraversion Item 1). Another child, age 6, said that in the picture with the map and globe, (Openness Item 3), the figure who didn't have the book was thinking. Another child age 7 suggested that the figure with the globe (Openness Item 3) wouldn't share it.

One child, age 9, said they felt confused by the scene of the group talking by the stairs (Extraversion Item 1).

All children reported preferring the animations to the static scenes.

Children age 10, reported enjoying the assessment and didn't raise any problems, nor did the adults.

Conclusions and thoughts about next steps

The early development work suggests that using images to assess personality traits is a promising method, including use with children under the age of 10 years but more development is needed. It should be noted that this first pilot was small scale and more data is needed to be able to make firm conclusions.

The patterns of responses for the younger age group were similar to the older participants but the younger children responded more positively to the Openness items. There were also marked differences for the items relating to neuroticism with the younger children being less neurotic for all but the Neuroticism 4 item; the swimming pool.

The low correlations between items intended to assess the same dimension needs to be further considered. In addition to needing more items in each scale, we need to carefully consider the item content. Our approach was to use figures within scenes but it may be that respondents are evaluating their own preference to the situation itself rather than their response being a reflection of the personality dimension. For example, a child may be fearful of swimming due to prior experience but not demonstrate neuroticism in other situations. Another consideration is within-person variation; we may feel and behave differently in different situations as we respond to expectations rather than our own preferences.

Thoughts about next steps

- Scale up the number of items in each dimension to 15 with a view to being able to create an assessment which separates out the dimensions.
- Change the colour of the figures within a scene so that they are all the same colour to avoid some children selecting a figure because of colour preference.
- Examine the facial expressions of the figures and consider making some of them more exaggerated to portray the emotion more clearly.
- Have the same number of figures which can be selected in each item.
- In creating new items each design should be run past a personality expert for blind categorization.
- Conduct further trialling with children under age 10.
- Extend the trialling to different groups, possibly in different countries.
- Consider using a rating scale and ask for each figure in a scene, how like the figure you feel.

References

Borghans, L., Duckworth, A.l., Heckman, J.J. and ter Weel, B. (2008, Fall). The economics and psychology of personality traits. *Journal of Human Resources 43* (4), 972–1059.

Carneiro, P., Crawford, C., Goodman, A. (2007). *The impact of early cognitive and non-cognitive skills on later outcomes. Pub. London: Centre for the Economics of Education*. ISBN 978-0-85328-188-7.

Costa, Jr., P.T. and McCrae, R. R. (1992) *NEO PI-R Professional Manual*, Psychological Assessment Resources, Inc, Odessa, FL.

Duckworth, A.L. and Yeager, D.S. (2015). Measurement Matters: Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes . Educational Researcher, Vol. 44 No. 4, pp. 237–251. DOI: 10.3102/0013189X15584327.

Easton, J. (2013). *Using measurement as leverage between developmental research and educational practice*. Paper presented at the Center for Advanced Study of Teaching and Learning Meeting, Charlottesville, VA. Retrieved from http://ies.ed.gov/director/pdf/Easton062013.pdf

John, O. P. and Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement and theoretical perspectives. In L. A. Pervin and O. P. John (Eds.), *Handbook of Personality: Theory and Research*, Chapter 4, pp. 102–138. New York: The Guilford Press.

Kautz,T., Heckman, J.J., Diris,R., ter Weel, B. and Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success*. Pub. OECD. https://www.oecd.org/edu/ceri/Fostering-and-Measuring-Skills-Improving-Cognitive-and-Non-Cognitive-Skills-to-Promote-Lifetime-Success.pdf

Merrell, C. and Tymms, P. (2016) Assessing Young Children: Problems and Solutions. In *Understanding What Woks in Oral Reading Assessments*. Montreal: UNESCO Institute for Statistics (UIS). ISBN 978-92-9189-196-2, Ref: UIS/2016/LO/TD/9, DOI: http://dx.doi.org/10.15220/978-92-9189-196-2-en

OECD (2015). *Skills for social progress: The power of social and emotional skills.* OECD Skills Studies, OECD Publishing. http:// dx.doi.org/10.1787/9789264226159-en

Paunonen, S.V. (2003). Big Five Factors of Personality and Replicated Predictions of Behavior. *Journal of Personality and Social Psychology*, 84 (2), 411–424.

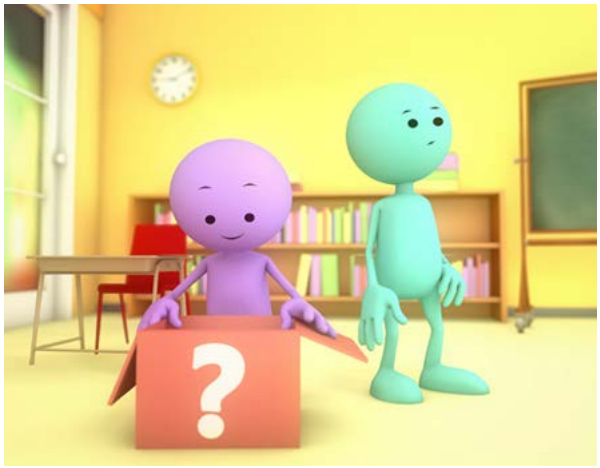
Paunonen, S.V., Trzebinski, J. and Forsterling, F. (1992). Personality Structure Across Cultures: A Multimethod Evaluation. *Journal of Personality and Social Psychology*. 62 (3), 447-456.

Paunonen, S.V., Jackson, D.N., Keinonen, M. (1990). The structured nonverbal assessment of personality. *Journal of Personality*, 58 (3), 481 – 502.

Soto et al. (2011) Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology*, 100 (2) 330 – 348.

Appendix

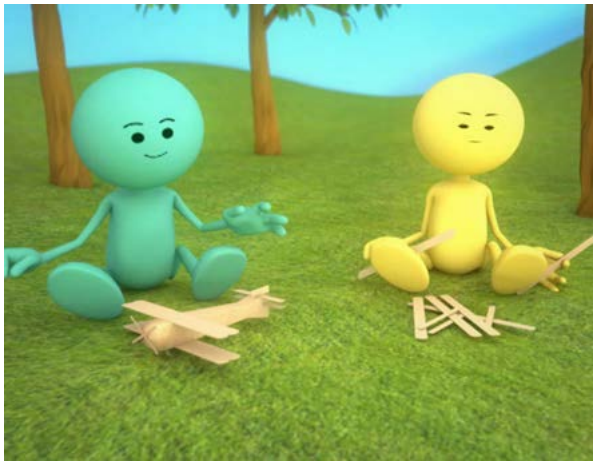
Openness Item 1



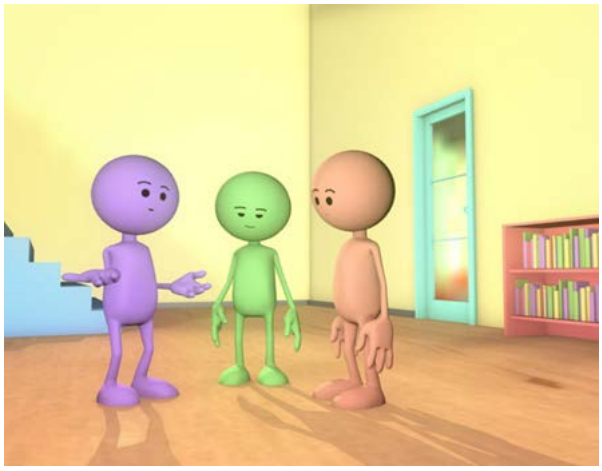
Openness Item 2 - Animated

Openness Item 3

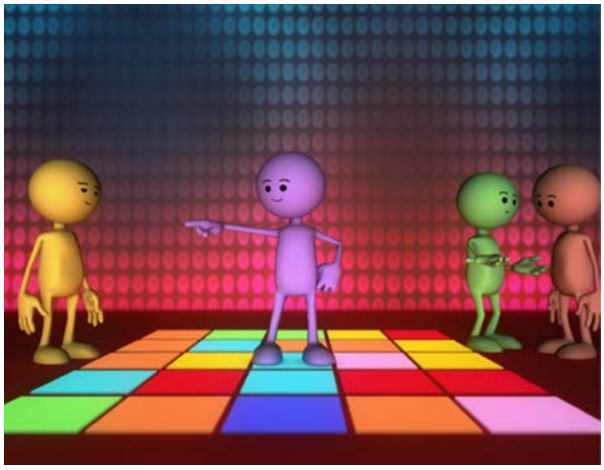
Openness Item 4 – Animated



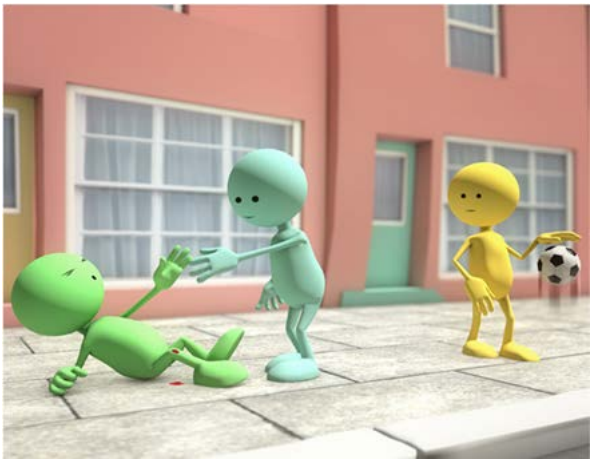
Extraversion Item 1 - Animated



Extraversion Item 2 – Animated

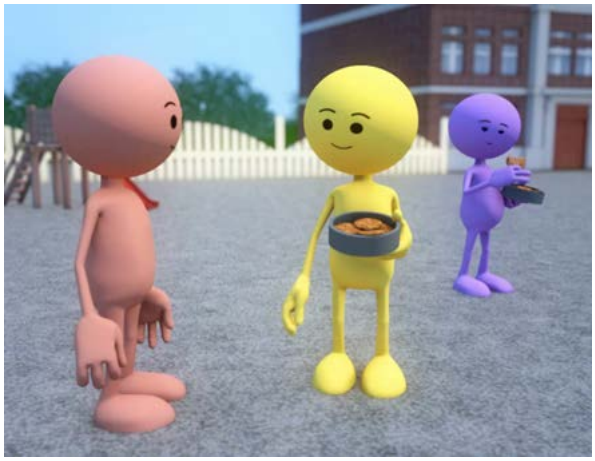


Agreeableness Item 1

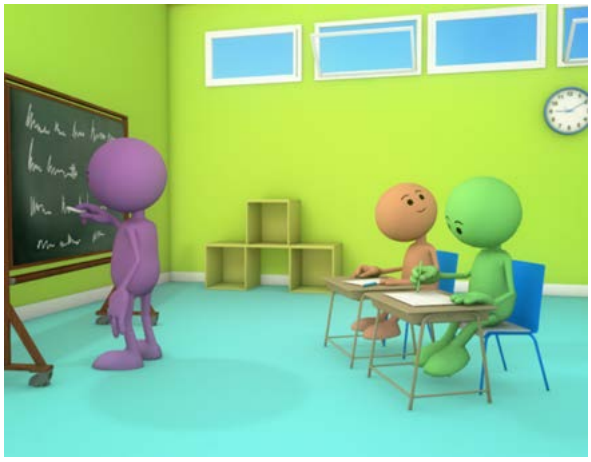


Agreeableness Item 2 - Animated

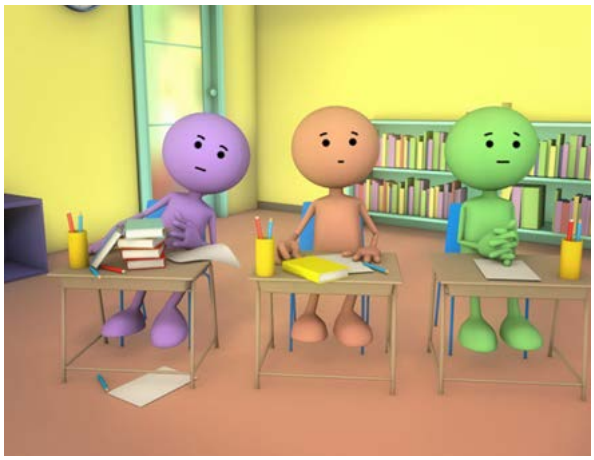
Agreeableness Item 3 – Animated



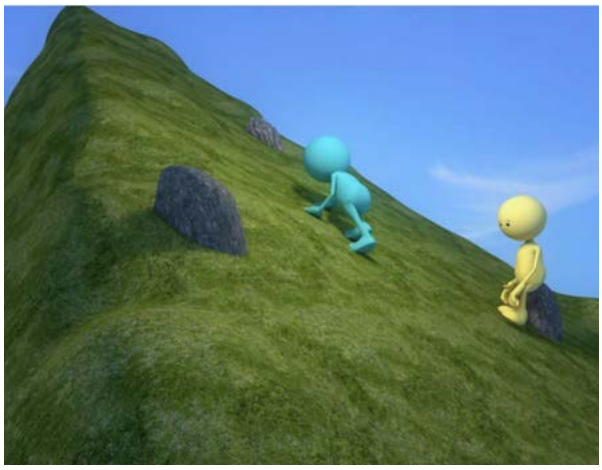
Conscientiousness Item 1 - Animated



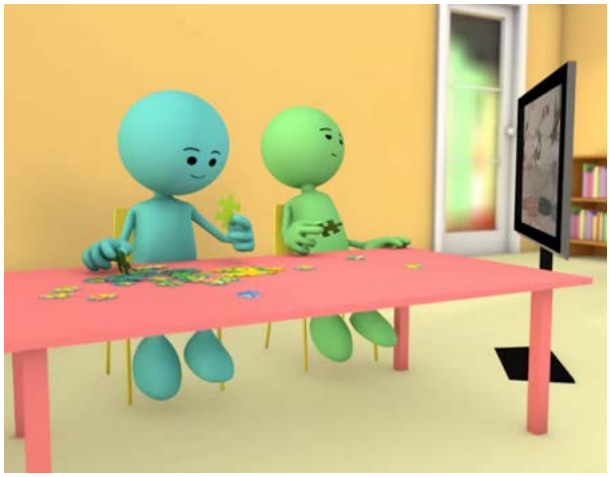
Conscientiousness Item 2



Conscientiousness Item 3 - Animated

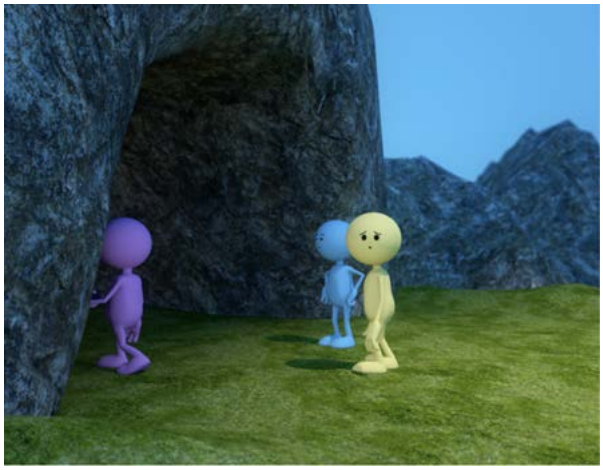


Conscientiousness Item 4 – Animated

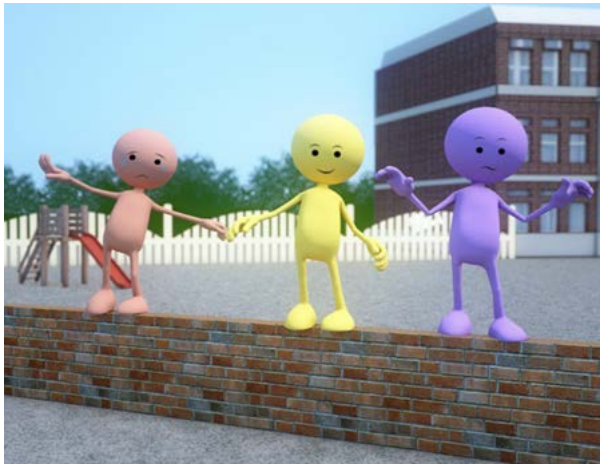


Neuroticism Item 1

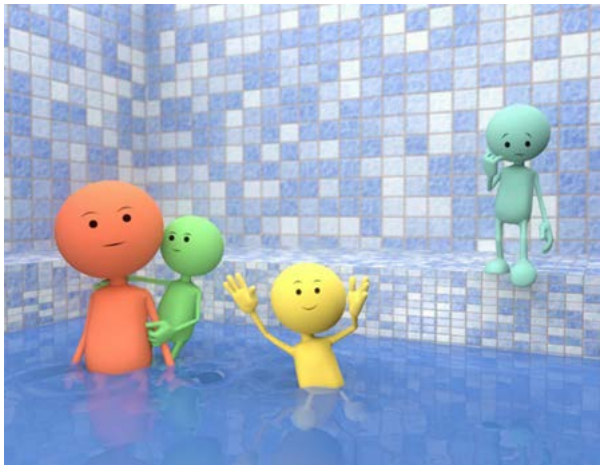
Neuroticism Item 2 - Animated



Neuroticism Item 3



Neuroticism Item 4



Neuroticism Item 5 – Animated