Construction and Validation of a Narrative Intelligence Scale with the Rasch Rating Scale Model

Reza Pishghadam
Ferdowsi University of Mashhad, Iran

Purya Baghaei
Islamic Azad University, Mashhad Branch, Iran

Mohammad Ali Shams
Ferdowsi University of Mashhad, Iran

Saeedeh Shamsaee
Ferdowsi University of Mashhad, Iran

Abstract

The aim of this study was to construct and validate a narrative intelligence scale. Randall (1999) defines narrative intelligence as the capacity both to compose and follow a story. The researchers developed a 35-item narrative intelligence scale based on the dynamics of NI proposed in Randall' framework including emplotment, characterization, narration, genre-ation and thematization. Rasch rating scale model (Andrich, 1978) was utilized to demonstrate the construct validity of the scale. Six items did not fit the model and six other items were found to be redundant. Reliability, person and item separation indices were high for the whole scale and a 5-point Likert scale functioned effectively after eliminating the misfitting items.

Keywords: Narrative intelligence, Validation, Rash rating scale model

Introduction

The ability to create and comprehend narrative patterns is a distinct feature of human mind (Fecica & O'Neill, 2010). Researchers in various fields are interested in finding the dynamics of narrative ability. Narrative skill, in most of the studies is taken as a developmental mental ability (Metaes & Sengers, 1998) that grows under the influence of genes (Randall, 1999), education, cultural atmosphere (Kintsch & Greene, 1978), social factors (Labov, 1981, 1997, 2001; Shiro, 2003), social interaction (Bakhtin, 1981), and aging (Randall, 1999). To the authors` knowledge, no test has been designed to date to measure narrative intelligence. Randall's (1999) theory of narrative intelligence was originally meant to provide gerontologists with a framework with more explanatory power for analyzing the variables associated with aging. According to Phoenix, Smith and Sparkes (2010), within aging studies, narrative has been widely used to examine social policy (Biggs, 2001), issues of masculinity (Smith, Braunack-Mayer, Wittert, & Warin, 2007), interpersonal and intimate relationships (Rosenfeld, 2003), physical activity in later life (Dionigi & O'Flynn, 2007; Tulle, 2007), narrative maps of aging (Phoenix & Sparkes, 2009), and carework (Ribeiro, Paú, & Nogueira, 2007).

In developmental psychology, narrative is thought of as a way in which humans make sense of the world. Bruner (1991) believes that narrative has a basic role in human understanding of intentional behavior. According to him,

humans make sense of intentional action by assimilating it to a narrative structure. Psychological studies dealing with identity issues make use of narrative analysis to investigate the representation of self in individuals' personal stories (e. g. cf. Ochberg, 1994). Bers (2002) elaborates on how identity questions like "who am I?" and "What is my place in the world?" are answered by the use of different types of narratives: personal stories, popular tales, and cultural myths. Literature in narrative psychology (Bruner, 1987, 1998) and developmental psychology (Scott, 1988; Nelson, 1993 cited in Dautenhahn, 2002) stress the central role of stories in the development of a social self in human beings, and their implications for AI systems.

Devising a scale for measuring people's narrative intelligence can contribute to various fields of study including the theory's birth field i.e. aging studies and other field who use narrative as a mode of inquiry (see Liu & Xu, 2011).

Randall's Theory of Narrative Intelligence

Randall (1999) proposes that narrative intelligence consists of intertwining sub-capacities as the ability to emplot, characterize, narrate, genre-ate, and thematize. Along with this artful borrowing from story-telling terminology, he makes use of some familiar terms related to movies, and some others that carry certain technical meanings with them.

Emplotment. In simple terms, plot is what happens in a story. Randall's (1999) Emplotment consists of *editing, summarizing, coping with conflicts, prioritizing, perceiving events as events, connecting events, comprehending, filling in the blanks,* and *generating alternatives.* Pavlenko (2007) sees plot as the main element that allows storytellers to create engaging and coherent stories. The actions reported in narrative give plot-advancing foreground information to the audience (Hopper & Thompson, 1980 cited in Minami, 2004). In Labov and Waletzky's (1967) framework, action depicts the sequence of specific, chronologically ordered events comprising the speaker's experience.

Characterization. In narrative we characterize both ourselves and others; we build "a working picture of what we are like" and "construct theories" on what type of person others are (Randall, 1999, p. 17). Labov (2001) states that narratives that center on conflict are normally concerned with the assignment of responsibility for these events and polarizing the characters; while the protagonist conforms to all community norms, the antagonist violates those norms (Labov, 2001). Some of the speculation made by Randall (1999) about the quality of our characterization, we think, may go beyond the desired borders of the objective measure we are looking for. For example, he maintains that appreciating one's novelty and avoiding prejudging or stereotyping people reflects the high quality of the characterization on the part of the narrator. This entails that to measure this narrative sub-capacity, one should first decide on the extent to which the narrator is fair and unbiased toward the characters in the narrative. The subjectivity associated with such decisions goes against the objectivity expected from the linguistic indicators in our checklist.

Narration. Narration is the central sub-ability and the main essence of narrative intelligence. The way we narrate determines the level of communication we maintain with others. By communication, Randall (1999) means the ability to convey the meaning of the story by the use of logical connectives which show the connection between causes and consequences. We organize events into a narrative structure and try to fit them to temporal mashes. It can be argued that grammatical complexity and the use of more sophisticated cohesive devices influence the perceived *quality* of a story. Sometimes we "vary our vocabulary to accommodate the capacities of our audience" (Randall, 1999, p. 18). According to him, a good narrator is the one who gives neither too much details nor too little; and also should properly distinguish between self as subject or knower and object or known. Applying a particular narrative tone (optimism/negativism; manner of conceptualization) is a part of our narrative ability.

Genre-ation. Each narrative includes particular happenings that are tokens of broader types or genres. Bruner (1991) considers this particularity as one of the defining features of narrative. He emphasizes the pattern-making role of different narrative genres and maintains that genres provide us with conventional models and frameworks for making sense of human happening. Randall (1999) believes that we perceive a particular chain of events by recalling the conventional characteristic of the genre to which the narrative belongs. He maintains that the ability to genre-ate enables us to organize events into predictable patterns, and to perceive our life in a dramatic shape. In this view, concepts such as "success and failure" or a "happy and tragic" life find their meaning under genre-ic headings.

Thematization. The main idea of a narrative is called theme. A good narrator can easily identify the theme from recurring patterns of meaning that are observed in particular situations. Identifying symbols that refer to a particular point is also another ability which is categorized as one of the components of thematization (Randall, 1999).

Method

Participants

A community sample of 101 people participated in this study, comprising 33 males and 68 females aged between 20 and 28. All of the participants were university students attending two universities in Iran (Mashhad and Kashan), majoring in English Language Literature. These students were accepted to participate voluntarily in the experiment.

Instrumentation

The research instruments include the NIS which was developed and validated for the purpose of this study, and two prompts for eliciting the participants' narrative performance in two separate narrative tasks: A 10-min segment of the movie "Defiance" (2009) for the narrative reconstruction task

(Task 1), and an elaborated question asking the participants to recount the memory of the first day of elementary school in the personal narrative task (Task 2). NIS consisted of 35 items measuring different aspects of narrative intelligence based on the framework proposed by Randall (1999). All of the devised items were annotated by further explanation in Persian and English, and clear examples that specifies the linguistic realization of the dynamics of narrative intelligence.

The movie which was used for narrative reconstruction rendered the first 10 minutes of the movie "Defiance." This prompt was chosen for several reason: (1) the first 10 minutes of the movie constitute a clear sub-story (episode) with specific beginning and end; (2) almost all of the events in this 10-min segment can be completely perceived by the audience without any need for referring to further comments or explanations by the characters; (3) since the story happens in a non-English setting, the English language used by the actors and actresses is easily understandable for even intermediate learners of English as a foreign language, let alone students of English Literature who are about to finish their undergraduate program; and (4) the content of the story is mainly conveyed through visual messages which are deliberately designed and arranged by the director rather than lengthy and confusing dialogues.

Task 2 (the personal narrative task) was organized based on the prompt "*please tell the story of your first day in the elementary school*". This prompt was also chosen for several reasons: (1) personal stories that are produced out of this narrative prompt are homogenous to an acceptable extent; (2) most of people do have a memory of such a memorable day; it is pretty common experience; (3) settings, characters, events, and temporal sequence of the event in the required story are predictable to a high degree because all of them are universally bound to social and educational norms in the country; and (4) since the first day of school is not an ordinary occasion, most people have a relatively solid image about it, and can recount the details much better than many other occasions in their personal life.

Procedure

To substantiate the content validity of the NIS, the authors, based on the guidelines provided by Randall (1999), designed a scale consisting of 35 items. To disambiguate the items and to ensure the content validity of the scale, few students were asked to watch a movie and narrate it. Then, two experts in narrative intelligence theory were asked to rate the individuals' performance. Each item received a score of 1 to 5.

Before administering tasks 1 and 2, the participants were fully informed about the procedure and rubrics of the test (e.g. they were not allowed to take notes; participant had to only rely on their cognitive and memory capacities to retell the story). The recording session started immediately after watching the movie; it should be mentioned there was a 3-minte break between Task 1 (narrative reconstruction) and Task 2 (personal narrative) so that the participants could have recollect and organize their memories and become prepared for performing the second task. In both tasks, the participants were required to tell the story in their mother-tongue (Persian). The narrative performance of the participants was rated by two raters using NIS. A high correlation (.91) between

the scores given to 101 participants of the study by the raters showed acceptable inter-rater reliability. The details of data analysis are presented in the following section.

Results

Validation

The subjects were rated on a five-point Likert scale on the 35 NI items, with higher scores indicating more narrative intelligence. Rasch rating scale model (RSM) (Andrich, 1978) as implemented in Winstpes was used to analyze the data. Fit statistics for the items, reliability and separation statistics as well as rating scale statistics and principal component analysis of residuals were studied to investigate the validity of the instrument.

The initial analysis of all 35 items yielded an itemseparation index of 12.78 with an item reliability of 0.99, and a person separation index of 8.04 with a person reliability of 0.98. Item measures ranged from -2.27 (distinguishing between main plot and subplots) to 11.81 logits (Mimicking the characters' tone or voice). The "root mean square error" (RMSE) for items is 0.21 and for persons is 0.32 which indicate quite precise measurement. The analysis revealed that severalitems did not fit with model expectations, which suggested that they might measure a construct that isdifferent from other items or represent another dimension for the construct of narrative intelligence.

Table 1 shows that six items grossly misfit the model following the criteria set by Wright and Linacre (1994) for rating scale data (infit*MNSQ*, 0.6 to 1.4). The misfitting items in descending order of infit*MNSQ* index are items 34, 18, 27, 11, 15 and 20. Misfitting items which are flagged by infit*MNSQ* indices greater than 1.4 are indicators of multidimensionality. A major concern of construct validity is the idea that all the items are operationalization of a single underlying construct. In other words, the instrument developer attempts to define and represent his theoretical construct in a set of items. Misfitting items with infit*MNSQ* indices greater than 1.4 are items which do not contribute to the practical representation of the theory of the intended construct.

Six items seem not to operate in concord with the other items. Two decisions can be made on the basis of this finding: either delete the six items as irrelevant to the construct of narrative intelligence or carefully study the content of these items, compare them with the other items and the theory of our construct to better understand why they do not accord with the rest of the items. These sorts of qualitative investigations can help the researcher form a deeper understanding of the construct and fix his theory of the construct. For instance, Item 34, which has the worst fit index is "identifying special signs", i.e., the ability of a narrator to mention certain objects, shapes, pictures, action... that signify a defined and specific meaning (like cross showing a commitment to Christianity). The second worst fitting item is Item 18 "Mimicking the characters' tone or voice" which focuses on the narrator's ability to changehis natural voice to show a certain quality associated with the voice of characters. It seems that these abilities are not part of the construct of narrative intelligence or if they are, they form another dimension or aspect of narrative intelligence which require their own score profile and scores on these items should be reported separately.

Table 1 also shows that there are six items with infit*MNSQ* indices smaller than 0.6. These items in ascending order of infit MNSQ are Items 31, 28, 21, 22, 24, and 12. Items with small *MNSQ*infit indices are not degrading to measurement (Linacre, 2007). They only indicate redundancy or local dependency in the data. In other words, these items are duplicating each other and do not add further information.

Table 1
Item Measures and Fit Statistics

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
18	14	100	11.81	.42	1.98	2.8	.86	.1	.55	.63	90.0	92.5	ITEM_18
34	35	100	9.46	.28	2.86	7.8	9.78	7.5	-.03	.71	65.0	82.2	ITEM_34
25	220	100	1.06	.17	.89	-.7	.75	-1.8	.88	.84	60.0	62.6	ITEM_25
19	226	100	.88	.17	1.21	1.4	1.17	1.1	.86	.84	50.0	61.6	ITEM_19
13	233	100	.68	.17	1.29	2.0	1.28	1.8	.88	.84	45.0	60.9	ITEM_13
14	233	100	.68	.17	1.19	1.3	1.14	1.0	.88	.84	49.0	60.9	ITEM_14
9	240	100	.48	.17	1.17	1.2	1.12	.9	.88	.84	48.0	59.8	ITEM_9
1	248	100	.26	.17	.68	-2.6	.66	-2.8	.88	.84	67.0	58.5	ITEM_1
15	255	100	.06	.17	1.53	3.4	1.53	3.4	.74	.84	41.0	57.7	ITEM_15
29	258	100	-.02	.16	.90	-.7	.90	-.7	.93	.84	48.0	57.4	ITEM_29
23	268	100	-.29	.16	.76	-1.8	.78	-1.7	.91	.84	54.0	57.2	ITEM_23
30	271	100	-.37	.16	1.37	2.5	1.35	2.3	.76	.84	56.0	57.6	ITEM_30
17	275	100	-.48	.16	1.23	1.6	1.17	1.2	.84	.84	56.0	58.9	ITEM_17
7	276	100	-.50	.16	1.17	1.2	1.27	1.9	.92	.84	50.0	58.7	ITEM_7
33	276	100	-.50	.16	1.19	1.4	1.18	1.3	.83	.84	46.0	58.7	ITEM_33
4	280	100	-.61	.16	.77	-1.8	.78	-1.6	.90	.84	62.0	58.9	ITEM_4
20	281	100	-.64	.16	1.51	3.2	1.32	2.1	.77	.84	62.0	59.0	ITEM_20
31	283	100	-.69	.16	.42	-5.3	.46	-4.8	.93	.84	71.0	59.2	ITEM_31
16	284	100	-.72	.16	.83	-1.3	.89	-.7	.88	.84	59.0	58.9	ITEM_16
26	285	100	-.75	.16	.83	-1.2	.73	-2.1	.83	.84	52.0	58.9	ITEM_26
21	286	100	-.77	.16	.47	-4.7	.56	-3.6	.93	.84	75.0	58.7	ITEM_21
22	288	100	-.83	.16	.53	-4.0	.52	-4.1	.87	.84	75.0	58.9	ITEM_22
10	289	100	-.85	.16	.61	-3.1	.58	-3.4	.85	.83	72.0	59.0	ITEM_10
2	294	100	-.99	.16	.82	-1.3	.87	-.9	.86	.83	58.0	60.2	ITEM_2
5	297	100	-1.07	.16	.71	-2.2	.82	-1.3	.84	.83	68.0	60.6	ITEM_5
6	298	100	-1.10	.17	.73	-2.0	.81	-1.4	.85	.83	65.0	60.7	ITEM_6
35	299	100	-1.13	.17	.72	-2.1	.67	-2.5	.82	.83	64.0	60.8	ITEM_35
32	306	100	-1.32	.17	.62	-3.0	.53	-3.8	.88	.83	73.0	61.4	ITEM_32
28	308	100	-1.37	.17	.45	-4.8	.49	-4.1	.86	.83	75.0	61.3	ITEM_28
12	310	100	-1.43	.17	.59	-3.3	.57	-3.3	.91	.83	73.0	61.4	ITEM_12
24	313	100	-1.51	.17	.57	-3.4	.47	-4.3	.88	.83	77.0	61.5	ITEM_24
3	315	100	-1.57	.17	1.37	2.3	1.37	2.2	.81	.83	48.0	61.6	ITEM_3
27	321	100	-1.73	.17	1.87	4.7	1.51	2.9	.69	.82	57.0	61.9	ITEM_27
11	326	100	-1.87	.17	1.57	3.3	1.72	3.7	.65	.82	51.0	62.0	ITEM_11
8	340	100	-2.27	.17	.84	-1.1	.78	-1.4	.84	.81	63.0	62.8	ITEM_8
MEAN	266.6	100.0	.00	.18	1.04	-.3	1.18	-.5			60.7	61.5	
S.D.	66.1	.0	2.75	.05	.51	3.0	1.51	2.8			11.2	6.7	

Items with small infit*MNSQ* also result in inflated reliability and give a fake impression of the precision of measurement. The high reliability reported here is to some extent the result of this kind of items.

Figure 1, item-person map, reveals that the items and their thresholds cover a wide range of the NI scale with the majority of thresholds clustering against the majority of persons. This shows that the instrument is well-targeted for the sample and the construct of NI is measured along a wide range of ability.
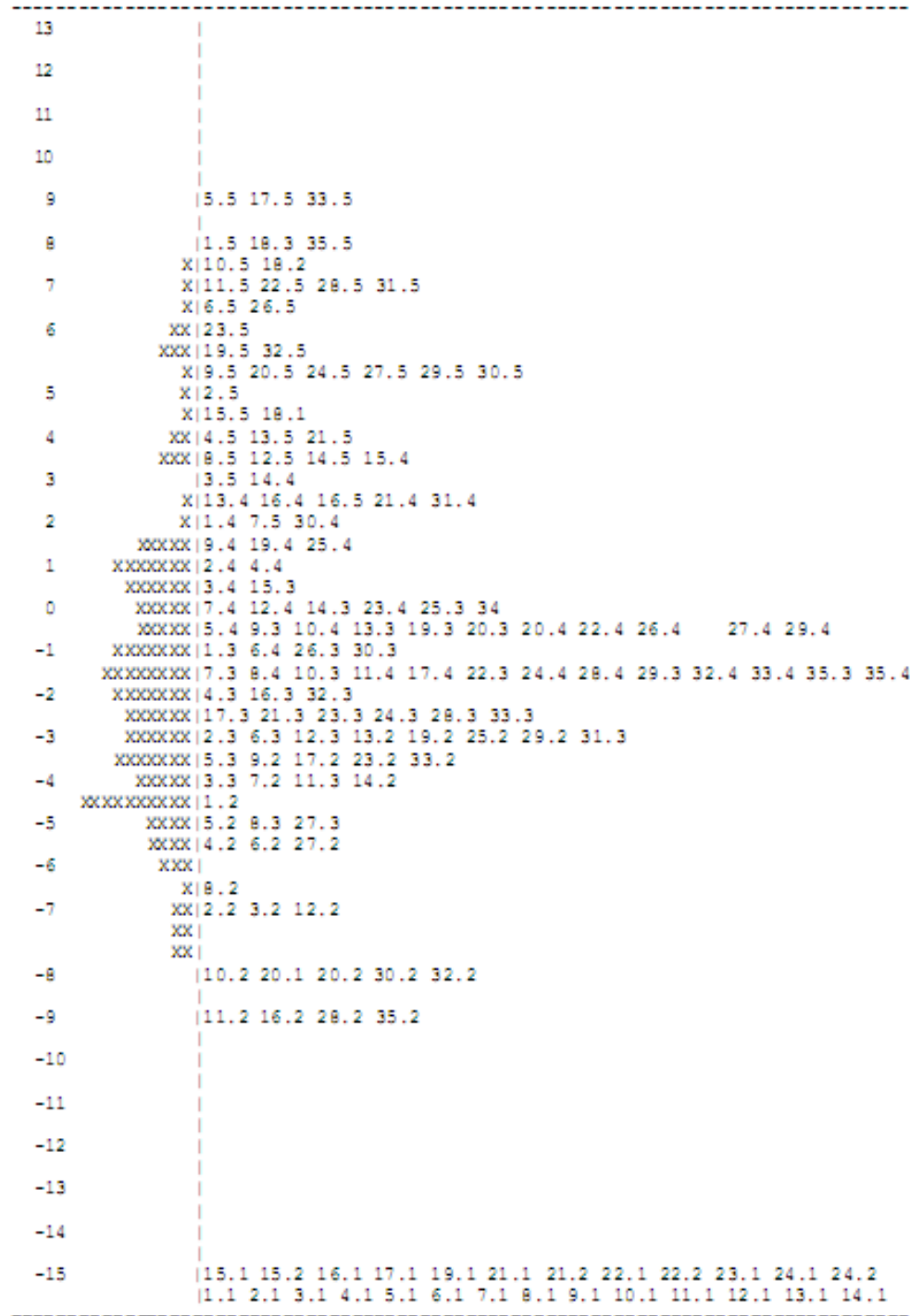
```
    ------------------------------------------------------------------------
  13                  |
                      |
  12                  |
                      |
  11                  |
                      |
  10                  |
                      |
   9                  |5.5 17.5 33.5
                      |
   8                  |1.5 18.3 35.5
                    X|10.5 18.2
   7                 X|11.5 22.5 28.5 31.5
                     X|6.5 26.5
   6                XX|23.5
                  XXX|19.5 32.5
                     X|9.5 20.5 24.5 27.5 29.5 30.5
   5                 X|2.5
                     X|15.5 18.1
   4                XX|4.5 13.5 21.5
                  XXX|8.5 12.5 14.5 15.4
   3                  |3.5 14.4
                     X|13.4 16.4 16.5 21.4 31.4
   2                 X|1.4 7.5 30.4
               XXXXX|9.4 19.4 25.4
   1       XXXXXXX|2.4 4.4
             XXXXX|3.4 15.3
   0          XXXXX|7.4 12.4 14.3 23.4 25.3 34
             XXXXX|5.4 9.3 10.4 13.3 19.3 20.3 20.4 22.4 26.4    27.4 29.4
  -1      XXXXXXX|1.3 6.4 26.3 30.3
       XXXXXXXX|7.3 8.4 10.3 11.4 17.4 22.3 24.4 28.4 29.3 32.4 33.4 35.3 35.4
  -2      XXXXXXX|4.3 16.3 32.3
             XXXXXX|17.3 21.3 23.3 24.3 28.3 33.3
  -3        XXXXXX|2.3 6.3 12.3 13.2 19.2 25.2 29.2 31.3
          XXXXXXX|5.3 9.2 17.2 23.2 33.2
  -4        XXXXX|3.3 7.2 11.3 14.2
      XXXXXXXXXX|1.2
  -5        XXXX|5.2 8.3 27.3
            XXX|4.2 6.2 27.2
  -6         XXX|
               X|8.2
  -7          XX|2.2 3.2 12.2
             XX|
             XX|
  -8            |10.2 20.1 20.2 30.2 32.2
                |
  -9            |11.2 16.2 28.2 35.2
                |
 -10            |
                |
 -11            |
                |
 -12            |
                |
 -13            |
                |
 -14            |
                |
 -15            |15.1 15.2 16.1 17.1 19.1 21.1 21.2 22.1 22.2 23.1 24.1 24.2
                |1.1 2.1 3.1 4.1 5.1 6.1 7.1 8.1 9.1 10.1 11.1 12.1 13.1 14.1
    ========================================================================
```

Figure 1. Map of persons, items and step difficulties.

Rating scale statistics as shown in Table 3 reveal that overall the rating scale performs well. As stated before, the subjects were rated on a five-point rating scale. "CategoryLabel" shows the label or the name that is given to each category and "Category Score" shows the numerical value which is given to each level of the category. "ObservedCount" shows the number of times each category is endorsed and "%" shows the percentage of this number. "ObsvdAvreg" is the average ability measure of the persons who have been rated on each category. We expect observed averages to increase with category values, which do here.

"SampleExpect" is the Rasch model expected average measures for the persons who have been rated on each category.

Table 2
Rating Scale Statistics

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	STRUCTURE CALIBRATN	CATEGORY MEASURE	
0	0	155	4	-10.30	-10.9	3.22	2.98	NONE	(-8.64)	0
1	1	458	13	-2.62	-2.21	.86	1.77	-7.54	-4.79	1
2	2	1096	31	-.22	-.32	1.04	.97	-2.03	-.59	2
3	3	719	21	1.34	1.44	.95	.93	.97	1.64	3
4	4	836	24	3.66	3.57	.92	.95	2.28	4.34	4
5	5	236	7	6.33	6.44	1.04	1.08	6.33	(7.44)	5

"InfitMNSQ" and "Outfit MNSQ" are the average of the infitand outfit mean-squares of the responses in each category. The expected value for these statistics is one. Gross deviations from one for these indices indicate unexpected observations. As it is shown in Table 3, the first category misfits. However, it seems that only a few very unexpected ratings have caused the misfit since the unexpectedness has not damaged the ordering of categories. When the six misfitting items are deleted the category fit improves too. Table 3 shows the category statistics after deleting the six misfitting items.

Table 3
Rating Scale Statistics After Removing the Misfitting Items

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	STRUCTURE CALIBRATN	CATEGORY MEASURE	
1	1	403	14	-4.89	-4.63	.71	.76	NONE	(-5.71)	1
2	2	937	32	-2.58	-2.75	1.09	.96	-4.59	-2.90	2
3	3	625	22	-.50	-.40	1.06	1.04	-1.16	-.28	3
4	4	728	25	2.17	2.08	1.00	1.05	.59	2.89	4
5	5	207	7	5.30	5.55	1.22	1.24	5.16	(6.27)	5

Structure calibration or Rasch-Andrich thresholds are the points on the rating scale where the probability of being observed in either of two adjacent categories is equal. The first category has no prior category so there is no measure for that. Threshold estimates show how difficult it is to observe each category. We expect threshold estimates to increase with category values. Disordering in threshold estimates, i.e., thresholds which do not advance with category values indicate that the category is rarely endorsed and has a narrow interval on the variable and the definition of categories are problematic (Linacre, 2007).

Steps which are labeled as say, "strongly agree", "agree", "neutral", "disagree" and "strongly disagree" should each define a different position on the construct of interest to be useful and meaningful. In fact, the raters or respondents should be able to distinguish the difference between these steps and easily associate different levels of the construct with each category. The number of categories should be small enough to be distinct and distinguishable for the respondent and raters and large enough to cover the entire range of the variable.

Threshold estimates show the distinctiveness of each step on a Likert scale. They also show the sufficiency of the number of categories. The thresholds should neither be too close nor too far from each other (Bond & Fox, 2007). Threshold values which are too far from each other show that the respondents can distinguish other steps or levels of the construct in between two categories and are forced to choose one because there is no other option. Suppose you ask you respondents to express their attitude towards some statements on a three-point scale, labeled as "very much", "to some extent", and "never". Too far apart threshold estimate on such a scale indicates that the investigator should increase the number of categories because the respondents are indeed capable of associating other levels of the construct with more steps on the rating scale. This will result in a more precise measurement of the construct.

Linacre (1999) recommend that the distance between thresholds should be at least 1.4 logits to indicate the distinctiveness of the steps and no more than 5 logits to avoid loss of information because of lack enough distinguishable categories. The distinction between thresholds can be investigated graphically by examining the probability curves. These curves show the probability of endorsing or being rated on a particular category of a Likert scale for every ability level. "Each category should have a distinct peak in the probability curve graph, illustrating that each is indeed the most probable response category for some portion of the measured variable" (Bond and Fox, 2007, p. 224). Disordered or too close categories which are problematic will be shown as having flat curves on the graph.

Figure 2 below shows the category probability curves for these data. As the graph shows each category, drawn with the number which indicates a given category, is most probable for persons with a given ability level on the variable. In fact, the threshold estimates are the intersections of rating scale category curves. "The plot should look like a range of hills. Categories which never emerge as peaks correspond to disordered Rasch-Andrich thresholds. These contradict the usual interpretation of categories as being a sequence of most likely outcomes" (Linacre, 2007, p. 241). Category measure is the difficulty of endorsing or being rated in each category. Again we expect an increase in these measures with category values.
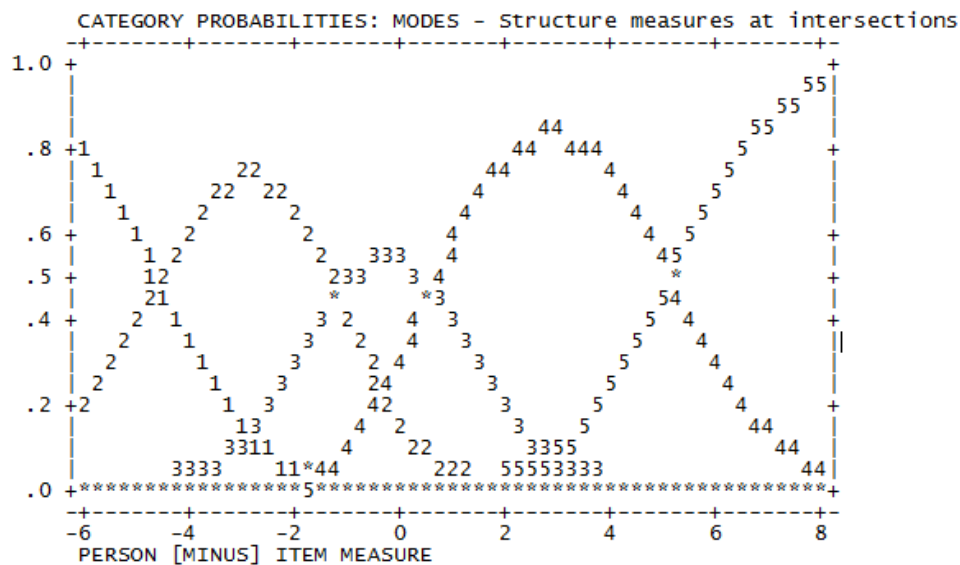
Figure 2. Category curves

As Figure 2 and Table 3 show the rating scale categories function properly, since structure calibrations, category measures and observed averages advance with category values and there is no disordering. Although the distances between structure calibrations is within the acceptable range of 1.4 to 5 logits, it seems that we can still add some steps to fill the gaps of around 4 logits in the lower and upper end of the scale.

Discussion and Conclusion

To measure narrative intelligence a 35-item inventory was designed. To design the inventory, the researchers went through a rigid process of validation. Rasch model was utilized to substantiate the construct validity of the inventory. The results of Rasch measurement showed that except for six items, all other items meet the unidimensionality criterion, laid down by the Rasch model. These items in descending order of infit*MNSQ* values are: mimicking the characters' tone or voice, identifying special signs, constructing various versions to account for specific events, conveying what is going on in the story clearly, perceiving a chain of events as tragic, comic, ironic and explaining events in terms of origins, outcomes, influences and results.

Moreover, it was found that six other items seem to be redundant, failing to add any further information regarding narrative intelligence, and can be deleted to increase the practicality of the inventory, resulting in a shorter instrument. Investigating the content of these six items confirms the statistical finding indicated by low infit*MNSQ* indices. Item 31 is "Using the conventional cultural shapes of life to describe situations" which is described as whether the narrator mentions the accepted cultural conventions such as faithfulness (husband and wife), obeying one's master, sacrificing oneself for one's country or belief. Now consider item 28: "Sensing the difference between a good mood and a bad mood" which relates to the narrator's ability to describe whether the reported events are bad or good in the eyes of the characters. As one can see, both items require making moral value judgments about reported events. Items 21 and 22

are "Using logical connectives between events" and" Using temporal junctures" respectively. These two items are defined as the narrator's ability to use expressions such as (because of, and this led to, A resulted in B) and the narrator' ability to use expressions such as (After that, then, before...) respectively. One can easily see the relationship between the two items. Items 24 is "Showing the importance of central actions in the story" i.e., whether the narrator mentions what actions are more important that others in the sense that they determine the chain of future events (if A had not done this, B would have not happened). Item 12 is "Making sense of events after the fact" or the ability of the narrator to refer back to some events that were merely reported (but not elaborated) earlier in the story, and his willingness to complete his account of them by mentioning more details and information. As can be seen here these two items are also very similar. Having a close look at the contents of the other items, one can acknowledge there are several other items with content similar to the contents of these six items.

It is our hope that future research will lead to further evaluation and improvement of this instrument. Researchers should continue to carry out thorough assessment of the psychometric properties of the instrument designed to measure narrative intelligence. Only after the true factor structure of the NIS has been examined, can researchers confidently assert conclusions about the role of this variable in education. The NIS designed in this study provides the researchers in various fields of study in social sciences with a valid measure of narrative ability of human subjects. This scale can be used along with different combinations of narrative tasks and research design to investigate the participants' narrative performance in different contexts.

References

Andrich, D. (1978). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement, 2,* 581-594.

Bakhtin, M. (1981). *The dialogic imagination.* TX: University of Texas Press.

Bers, M. U. (2002). We are what we tell: Designing narrative environments for children. In M. Mateas & P. Sengers (Eds.), *Narrative intelligence: Advances in consciousness research* (pp. 113-128). Philadelphia: John Benjamins.

Biggs, S. (2001). Toward critical narrativity: Stories of aging in contemporary social policy. *Journal of Aging Studies, 15,* 303–316.

Bond, T. G., & Fox, C. M. (2007) *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). London: Lawrence Erlbaum.

Bruner, J. (1987). Life as narrative. *Social Research, 54,* 11–32.

Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry, 18,* 1-21.

Bruner, J. (1998). What is a narrative fact? *Annals of the American Academy of Political and Social Science, 560,* 17-27.

Dautenhahn, K. (2002). The narrative intelligence hypothesis: In search of the transactional format of narratives in humans and other social animals. In

M. Beynon, C. L. Nehaniv, & K. Dautenhahn (Eds.), *CT 2001* (pp. 248-266). Berlin: Springer-Verlag.

Dionigi, R., & O'Flynn, G. (2007). Performance discourses and old age: What does it mean to be an older athlete? *Sociology of Sport Journal, 24,* 359–377.

Fecica, A. M., & O'Neill, D. K. (2010). A step at a time: Preliterate children's simulation of narrative movement during story comprehension. *Cognition, 116,* 368-281.

Kintsch, W., & Greene, E. (1978). The role of culture-specific schemata in the comprehension and recall of stories. *Discourse Processes, 1,* 1-13.

Labov, W. (1981). Speech actions and reactions in personal narrative. In D. Tannen (Ed.), *Analyzing discourse: Text and talk* (pp. 217-247). Washington DC: Georgetown University Press.

Labov, W. (1997). Some further steps in narrative analysis. *The Journal of Narrative and Life History, 7,* 395-415.

Labov, W. (2001). Uncovering the event structure of narrative. *To appear in Georgetown University Round Table.*

Labov, W., & Waletzky, J. (1967). Narrative analysis. In J. Helm (Ed.), *Essays on the verbal and visual arts* (pp. 12-44). Seattle: University of Washington Press.

Linacre, J. M. (1999) Investigating rating scale category utility. *Journal of Outcome Measurement, 3*(2), 103-122.

Linacre, J. M. (2007) *A user's guide to WINSTEPS-MINISTEP: Rasch-model computer programs.* Chicago, IL: winsteps.com.

Liu, Y., & Xu, Y. (2011). Inclusion or exclusion?: A narrative inquiry of a language teacher's identity experience in the 'new work order' of competing pedagogies. *Teaching and Teacher Education, 27,* 589-597.

Mateas, M., & Sengers, P. (1998). Narrative intelligence. In M. Mateas & P. Sengers (Eds.), *Narrative intelligence: Advances in consciousness research* (pp. 1-26). Philadelphia: John Benjamins.

Minami, M. (2004). The development of narrative discourse patterns in the learning of Japanese as a foreign language.

Ochberg, R. L. (1994). Life stories and storied lives. In A. Lieblich & R. Josselson (Eds.), *Exploring identity and gender: The narrative study of lives* (pp. 113-144). London: Sage Publication.

Pavelnko, A. (2007). Autobiographic narratives as data in applied linguistics. *Applied Linguistics, 28,* 163–188.

Phoenix, C., & Sparkes, A. C. (2009). Being Fred: Big stories, small stories and the accomplishment of a positive ageing identity. *Qualitative Research, 9,* 219–236.

Phoenix, C., Smith, B., & Sparkes, A. C. (2010). Narrative analysis in aging studies: A typology for consideration. *Journal of Aging Studies, 24,* 1-11.

Randall, W. L. (1999). Narrative intelligence and the novelty of our lives. *Journal of Aging Studies, 13,* 11-28.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research, 1960. (Expanded edition, Chicago: The university of Chicago Press, 1980).

Ribeiro, O., Paúl, C., & Nogueira, C. (2007). Real men, real husbands: Caregiving and masculinities in later life. *Journal of Aging Studies, 21*, 302–313.

Rosenfeld, D. (2003). Identity careers of older gay men and lesbians. In J. A. Holstein (Ed.), *Ways of aging* (pp. 160–181). Oxford: Blackwell Publishing.

Scott, C. (1988). A perspective on the evaluation of school children's narratives. *Language, Speech and Hearing in the Schools, 19,* 67-82.

Shiro, M. (2003). Genre and evaluation in narrative development. *Child Language, 30,* 165-195.

Smith, J. A., Braunack-Mayer, A., Wittert, G., & Warin, M. (2007). I've been independent for so damn long!: Independence, masculinity and aging in a help seeking context. *Journal of Aging Studies, 21*, 325–335.

Tulle, E. (2007). Running to run: embodiment, structure and agency among veteran elite athletes. *Sociology, 41*, 329–346.

Appendix
Narrative Intelligence Scale (NIS): 35 Items

	Item	*Realization in the Narration*
1	Explaining settings by referring to the 5 senses	the narrator describes the environment by mentioning certain sounds, smells, scenes, tastes, or textures
2	Good recalling of events, people, and settings	The narrator has a powerful memory and remembers the details of the story very well
3	Adequate use of summarizers	Enough number of expressions such as "finally", "at the end", etc. are used.
4	Appropriate use of summarizers	Summarizers are used at the right time and right place
5	Recognizing the conflicts	The narrator describes the conflicts between the characters (fights, arguments, physical challenges...)
6	Identifying the possible sources of conflicts	The narrator tries to explains why there is a conflict between the characters
7	Attempting a measure of resolution for conflicts	The narrator mentions the possible solutions through which the characters may find a way to finish the conflicts
8	Distinguishing between main plot and subplots	The narrator explicitly mentions that what pars of the story contribute to the main narrative and what parts can be considered as mini-narratives that only elaborate on one the aspects of the main narrative. The narrator does not digress by allocating much time to peripheral narratives while the main narrative is left unfinished.
9	Perceiving situations as discrete temporal units with beginnings, middles, and ends	Each event or scene is seen and described by the narrator as a temporal unit that begins at a certain point of time (at first...), unfolds in space and time (then..., next..., after that...), and finishes at another specific point in time (finally, at the end...)
10	Linking events in a consequential order	Each event causes the subsequent ones. The narrator orders the events according to their actual precedence in the story (what happens first is told first)
11	Explaining events in terms of origins, outcomes, influences and results	Narrators mentions whether the reported events are cause, effect, results of some other actions, or set the ground for some future actions
12	Making sense of events after the fact	The narrator refers back to some events that were merely reported (but not elaborated) earlier in the story, and tries to complete his account of them by mentioning more details and information
13	Envisioning events before they occur	The narrators predicts the occurrence of the future events in the story
14	Maintaining central story-lines	The narrator sticks to the main line of the story and does not bore the audience by explaining about aspects of the characters and events that are irrelevant to the main plot
15	Constructing various versions to account for specific events	The narrator tries to explains why something is happening the way it is happening, trying to provide a story that makes sense
16	To see situations from different angles	The narrator mentions a single event from different angels (from the view point of different characters) while trying to give a multi-faceted account of the reported event (e.g. to character A it was a disaster but for B, there was no problem...)
17	Imagining the characters' thoughts and feelings	The narrator tries to mention what is going on inside the head of characters (e.g. he was thinking, she decided to...)
18	Mimicking the characters' tone or voice	The narrator changes their natural voice to show a certain quality associated with the voice of chracters
19	Feeling compassions for the characters	The narrator makes sympathetic or empathetic comments (she was suffering and I could feel it, I felt bad for...)

	Item	*Realization in the Narration*
20	Conveying what is going on in the story clearly	The narrator mentions specific objects, and delineates the physical and mental atmosphere (by referring to the locations, coordinates, order, and size of objects and places)
21	Using logical connectives between events	The narrators uses expressions such as (because of, and this led to, A resulted in B)
22	Using temporal junctures	The narrator uses expressions such as (After that, then, before...)
23	Using neither too much detail nor too little	The narrator mentions the details of the story only when it is really needed (to characterize the characters, or to explain the settings)
24	Showing the importance of central actions in the story	The narrator mentions what actions are more important that others in the sense that they determine the chain of future events (if A had not done this, B would have not happen)
25	Using rhetoric moves to sustain the interest of the audience	The narrator does not tell all the interesting points from the beginning, sometimes he mentions something briefly and then goes to elaborate on it little by little by revealing pieces of information gradually hence maintain the interest of the audience and keeping them attentive to the rest of the story
26	Maintaining a particular tone (e.g. optimism, negativism, realism)	The narrator is either realist, or optimist, or pessimist while reporting the events. He is consistent in representing his attitude toward the constructed reality in the story
27	Perceiving a chain of events as tragic, comic, ironic	The narrators describes the series of events as tragic, ridiculous, etc.
28	Sensing the difference between a good mood and a bad mood	The narrators describes whether the reported events are bad or good in the eyes of the characters
29	Mentioning ups and downs of the lives described in the narrative	The narrator explicitly mentions that event B makes character A's life harder or easier
30	Imagining a dramatic shape for the events in the narrative	The narrator looks at each event as a factor that makes the situation tragic, comic, etc for the characters
31	Using the conventional cultural shapes of life to describe situations	The narrators mentions the accepted cultural conventions such as faithfulness (husband and wife), obeying one's master, sacrificing oneself for one's country or belief
32	Mentioning recurrent patterns in events	The narrator mentions that character A has done this or that several times; event B has happened frequently; setting C has been observed a number of times
33	Analyzing recurrent patterns in events (going to the details and describing their development)	The narrator tries to explain what some actions, events, or settings are more frequent that others (trying to provide a unifying explanations for apparently different and unrelated elements in the story)
34	Identifying special signs	The narrator mentions certain objects, shapes, pictures, action... that signify a defined and specific meaning (like cross showing a commitment to Christianity)
35	Finding particular points in events and comments	The narrator explain about the messages embedded in the story (ethical messages, journalistic implications, religious beliefs, political propaganda, personal perceptions, or director's intention to show or represent something in a certain manner)

Acknowledgement

About the Authors

Reza Pishghadam is associate professor in TEFL in Ferdowsi University of Mashhad, Iran. His research interests are: Psychology of language education and Sociology of language. Email: address is: pishghadam@um.ac.ir; rpishghadam@gmail.com

Purya Baghaei is an assistant professor in the English Department of Islamic Azad University, Ostad Yusofi St., Mashhad, Iran. His major research interests are the applications of unidimensional and multidimensional Rasch models in educational research. Email: pbaghaei@mshdiau.ac.ir.

Mohammad Ali Shams is a Ph.D. candidate of TEFL in Ferdowsi University of Mashhad, Iran. His major research interests include: Testing and Psycholinguistics. Email: neoshams@yahoo.com

Saeedeh Shamasaee is a Ph.D. candidate of TEFL in Ferdowsi University of Mashhad, Iran. Her major research interests include: Linguistics and Psycholinguistics. Email: shamsaee_student@yahoo.com