



The Tried, the True, and the Transpiring in Creativity Assessment

Patrick Kyllonen

Center for New Constructs (R&D)

Educational Testing Service

"CAN CREATIVITY BE MEASURED?"

28-29 May 2009

Brussels, MADOU auditorium

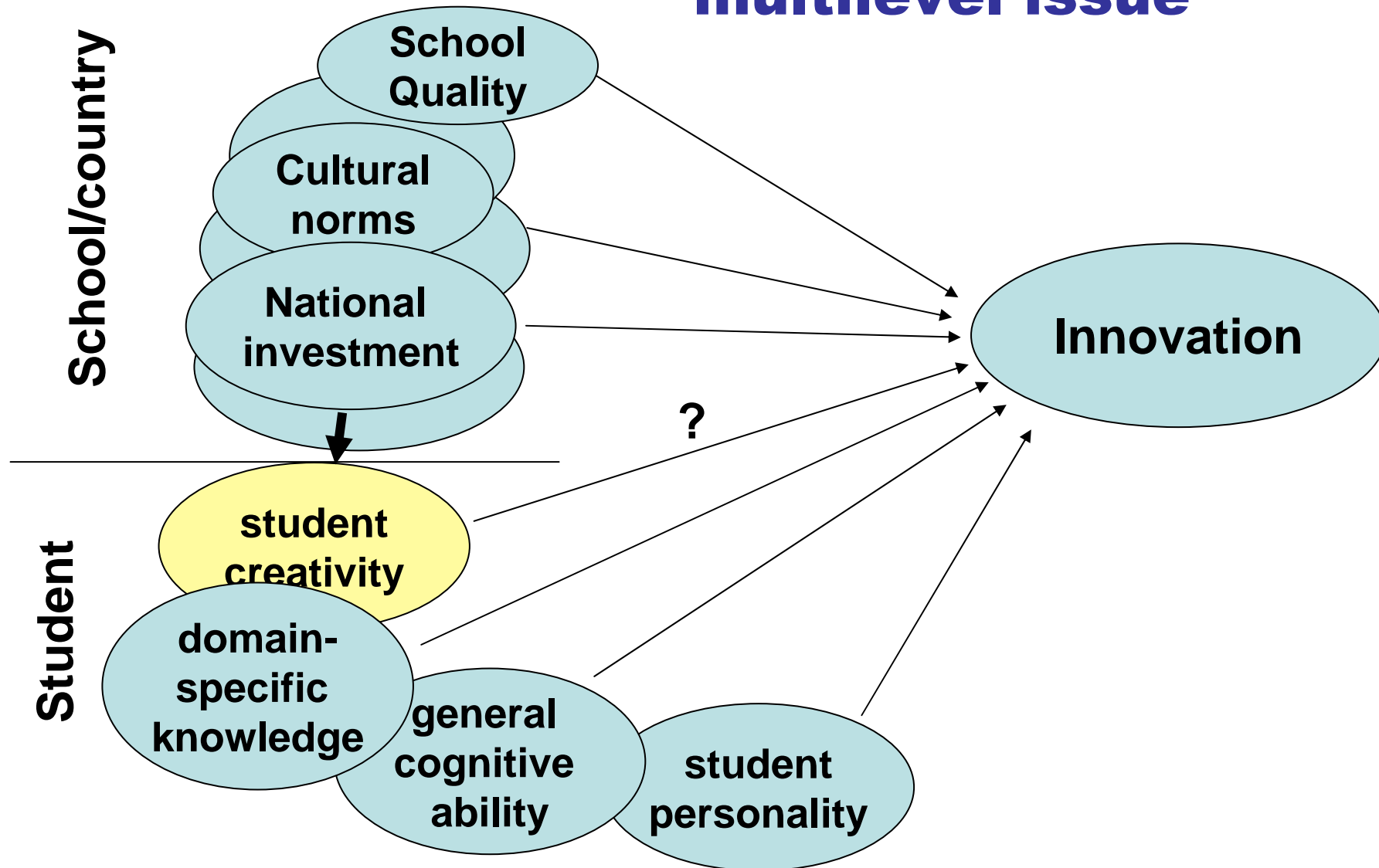
**Session 2 : Approaches to the
measurement of creativity at the
individual level (May 29, 1300—1430)**

Kyllonen, P. C. (2009). The tried, the true, and the transpiring in creativity assessment. In E. Villalba (organizer), *Can creativity be measured?..* OECD: Brussels, Belgium.

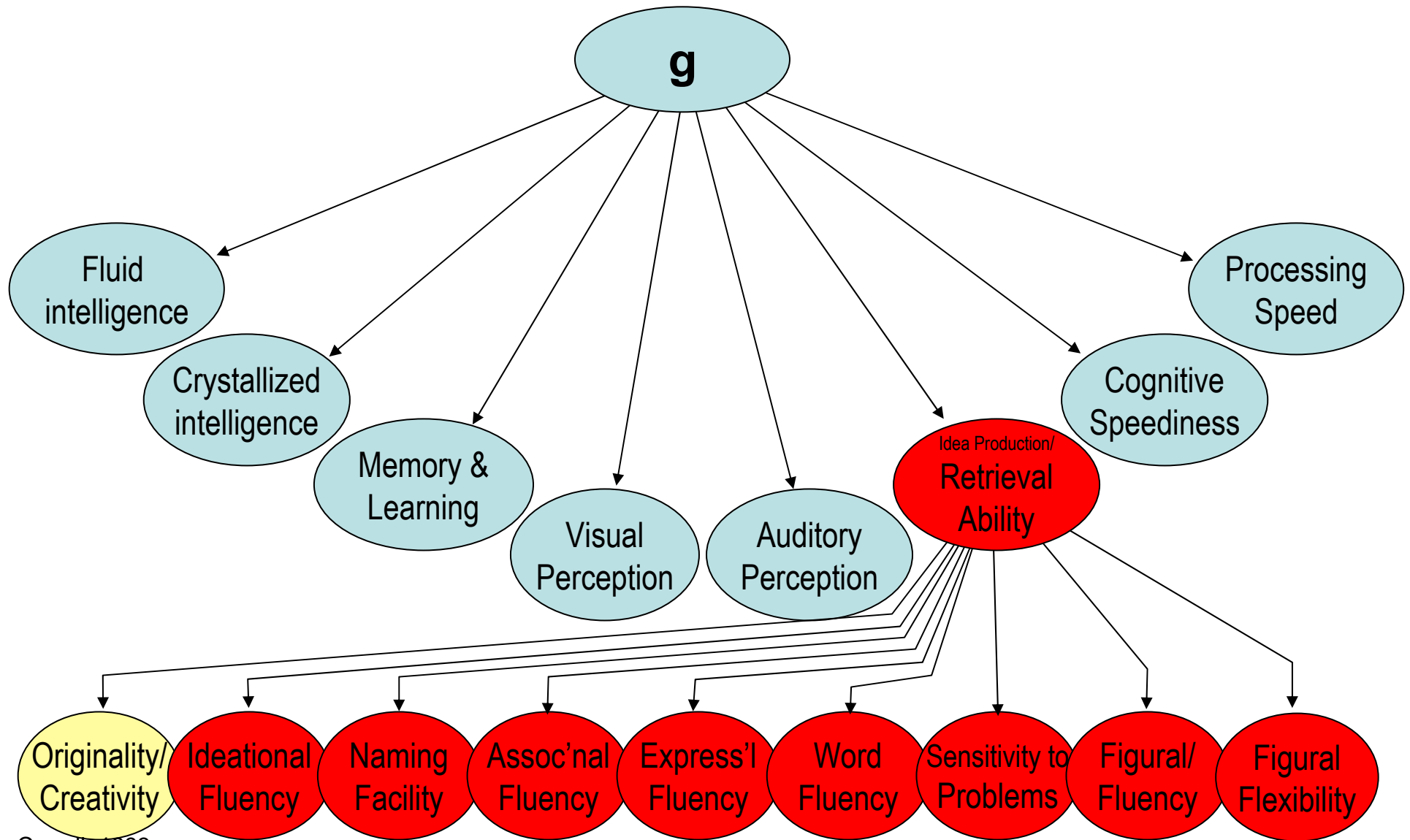
Outline

- Factor analytic investigations suggest what creativity is, and is not
 - Carroll's (1993) taxonomy of cognitive tasks
- ETS's "Formulating Hypotheses" test satisfies the "factor-analytic" definition of creativity
 - Fluency task
 - Difficult to generate many answers
 - Correlated with some creativity outcomes
- There is a process model account of creative problem solving that accommodates the factor analytic results (i.e., the centrality of fluency, but the idea that there is also something more)
 - Forbus & Gentner's "Many are called, few are chosen" model
 - Fluency stage + structural mapping (analogical reasoning) stage
 - Implications for instruction—encourage interdisciplinary analogies

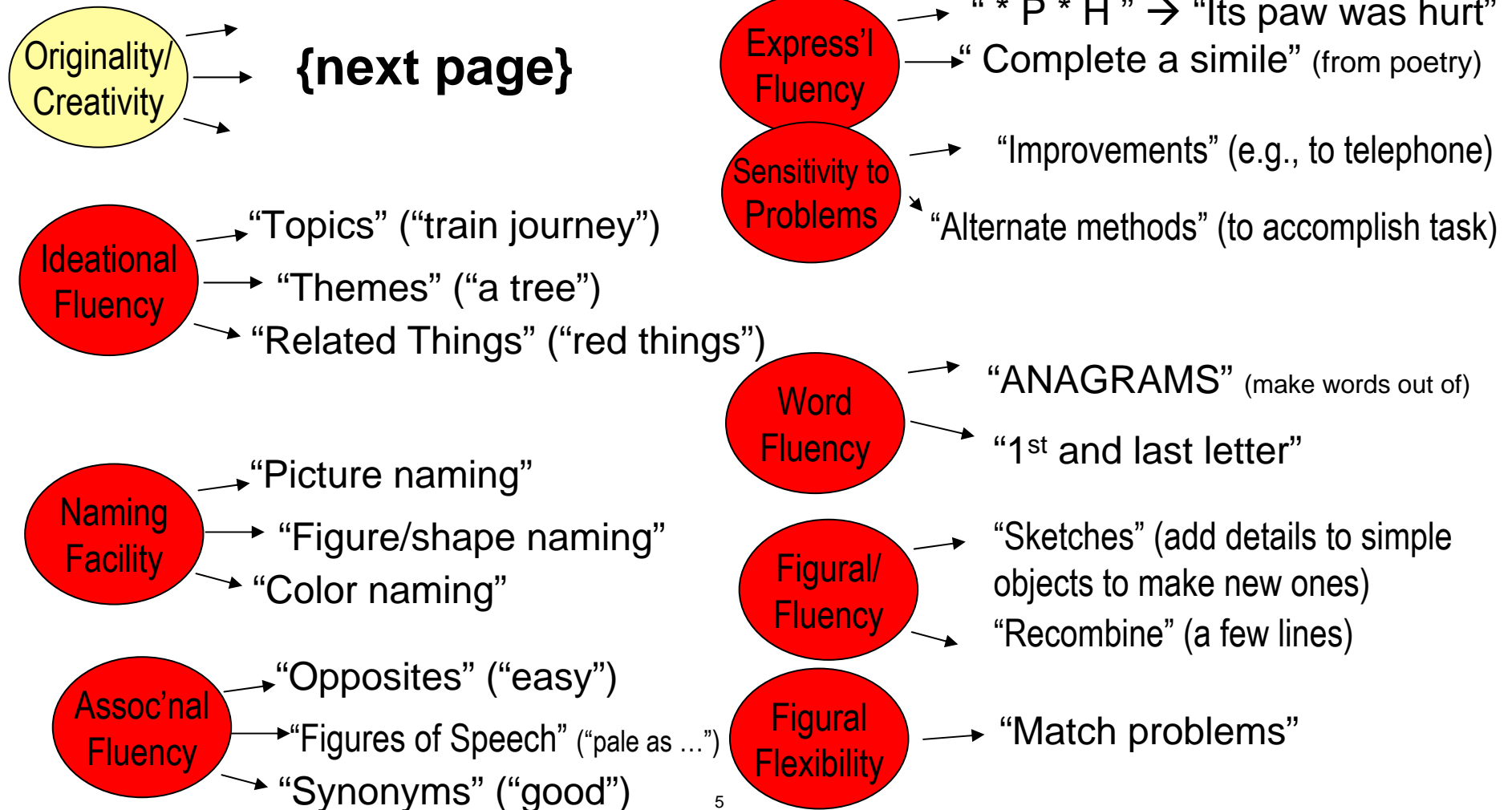
Determinants of Innovation is a multilevel issue



Carroll's 3-stratum Theory



Measures of Idea Production Factors



Originality/Creativity

- Types of tasks
 - Unusual uses of common objects (e.g., bricks)
 - Titles/Captions of stories/cartoons (e.g., Tufts' creativity test)
 - Consequences of hypothetical events (e.g., no sleep)
 - Grouping common objects by various rules
 - Picture meanings/stories (e.g., Kaufman's dissertation)
 - Other (anything requiring thinking of non-obvious responses)
- Scoring
 - Fluency: # responses
 - Flexibility: # category switches
 - Originality: via ratings or rarity
 - Elaboration: amount of detail

The Creativity Factor

- “The common element [in good measures of the creativity factor] is that they require examinees fairly quickly to think of...a series of responses fitting the requirements of the task...furthermore, ... it is difficult and challenging to think of responses beyond the more obvious commonsense ones” (Carroll, 1993; p. 428)
- I.e., Fluency + some additional process

ETS Scientific Creativity Tasks

- Based on Flanagan's (1949) "critical incidents" studies of research jobs; of 8 tasks, 3 were scientific thinking tasks
 - Formulating hypotheses; planning investigations; interpreting findings
- Four "Tests of Scientific Thinking" (Frederiksen & Ward, 1978)
 - Formulating Hypotheses (FH)
Write hypotheses to explain a finding ($M = 2.5, .6, .2$)
 - Evaluating Proposals (EP)
Critique the design, methodology, or theoretical position of student proposals ($M = 3.9, 1.1, .2$)
 - Solving Methodological Problems (SMP)
Write solutions to methodological problems encountered by graduate students ($M = 2.4, .5, .1$)
 - Measuring Constructs (MC)
Suggest methods to measure constructs (e.g., conservatism; bigotry; leadership) without ratings or self-reports ($M = 2.8, .8, .2$)
(M = Mean number of responses, unusual responses, unusual high-quality responses)

Darlington Playground

The playground at the Darlington Middle School has a swing set and a seesaw. Beginning in September and continuing through the year, this equipment is in almost constant use during the playground activity periods on each school day. Although school is open on March 4, none of the children are using either the swing set or the seesaw.

Think of hypotheses (possible explanations) to account for the lack of use of the equipment.

Write each hypothesis as a separate answer of no more than **7 words**.

1. **downed live power line on playground**
2. **tornado destroyed equipment**
3. **everyone is watching a fight**
4. **everyone has the flu**
5. **it is very cold that day**
6. **children are being punished**

7.

8.

9.

10.

11.

12.

13.

14.

15.

Edit

Save

the school lost its liability insurance

Test



Quit

Time



Help

Answer

Confirm



Next

Scoring rubric for Formulating Hypotheses

- Response categories predetermined
 - Categories are already rated (for quality—“most likely to be correct”)
 - Rater only has to categorize, not rate
 - Enables completely automated scoring
- Several scores (4 judges, high agreement)
 - Mean quality (of all responses) (.80)
 - Highest quality (of all responses) (.83)
 - Quality of response subject marked “best” (.80)
 - # responses (.90)
 - # unusual responses (.72)
 - # unusual, high-quality responses (.69)

(factor analysis shows a fluency factor & two quality factors, 1 for FH & MC; the other for EP & SMP)

1. SWING AND SEESAW EQUIPMENT NOT USED

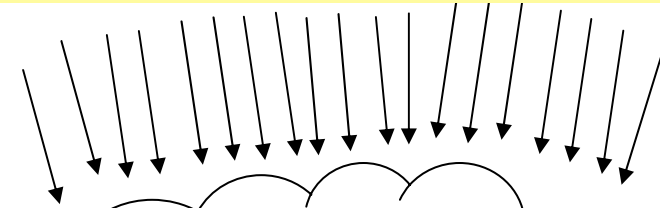
GENERAL CATEGORY	SPECIFIC CATEGORY
A) CHILDREN NOT AT SCHOOL	1. PLAYED HOOKY 2. TEACHER STRIKE 3. TEACHER CONFERENCE 4. EMERGENCY/DISASTER 5. AWAY ON FIELD TRIP, ON TOUR, ETC.
B) NOT ALLOWED TO PLAY	1. BEING PUNISHED 2. REQUIRED TO ATTEND ASSEMBLY, SPECIAL EVENT, SPECIAL VISITOR 3. RELIGIOUS OBSERVANCE - NO PLAYING 4. BEING TESTED, TRYOUTS, REHEARSALS 5. MUST NOT GET DIRTY OR MESSY, PICTURE DAY 6. BOMB SCAPE / FIRE DRILL
C) PREOCCUPIED ELSEWHERE	1. PLAYING A GAME, ON OTHER EQUIPMENT, ETC. 2. PARADE - ACCIDENT - EQUIPME, ETC., DISTRACTS THEM 3. SOMETHING NEW, TOY, PET, ETC., ENGAGES THEM
D) CONDITION OF CHILDREN	1. TOO TIRED TO PLAY 2. HURT, INJURED 3. ILL
E) CHILDREN'S FEELINGS / ATTITUDE	1. VERY SAD, UPSET OVER A DEATH, TRAGEDY, ETC. 2. BORED, DON'T FEEL LIKE PLAYING 3. AFRAID TO PLAY - SOMEONE HURT 4. REFUSE TO PLAY - BOYCOTT
F) NO ACTIVITY PERIOD TODAY	1. EARLY DISMISSAL 2. ANOTHER EVENT RAN INTO OVERTIME
G) PROBLEM WITH TEACHERS / ADULTS	1. LACK OF SUPERVISION 2. TEACHERS FORGOT ACTIVITY PERIOD 3. ADULTS USING THE EQUIPMENT
H) CONDITION OF SWING / SEESAW	1. OFF LIMITS, BEING INSPECTED, RECENT ACCIDENT 2. DANGEROUS (SPUNTERS, SHARP EDGES) 3. BROKEN 4. BEING REPAIRED 5. BEING PAINTED 6. REMOVED - LOANED 7. STOLEN 8. KNOCKED OVER BY STORM 9. NASTY GRAFFITI
I) CONDITION ON / NEAR PLAYGROUND	1. PUDDLES, MUD, TOO WET, SNOW, ICE ON EQUIPMENT 2. BEING PAVED, SEEDED 3. DOWNED POWER LINE 4. DANGEROUS PEOPLE (CROOKS, MOLESTER, DRUG DEALER) 5. DANGEROUS ANIMALS (BEEES, MAD DOG) 6. FIRE OUTSIDE 7. GATE LOCKED, NO KEY 8. BROKEN GLASS, ETC. 9. UNPLEASANT SMELL, SIGHT (VOMIT, DEAD ANIMAL)
J) BAD WEATHER	1. RAIN, LIGHTNING, SNOW, WINDS, COLD, HEAT
K) ECONOMIC REASONS	1. LIABILITY INSURANCE EXPIRED
L) OBSERVATION ERRORS	1. OBSERVATION WAS MADE TOO EARLY/NOT ACTIVITY PERIOD NOW 2. CHILDREN ARE BETWEEN TURNS ON THE EQUIPMENT

Results

- FW {
 - Fluency & quality scores form separate factors
 - Fluency = # of hypotheses, # unusualness hypotheses
 - Quality = Highest and mean quality
 - Quality scores correlate moderately with GRE V & Q
 - Higher validities for fluency scores than for GRE
 - In self-appraisal of experimental design, # of professional activities, # meetings, # publications, collaborative research
- BR {
 - Replicated FW's validity findings
 - 15-word constraint worked well (slightly better than 7)
 - Students liked the “Formulating Hypotheses” test

MAC/ FAC theory

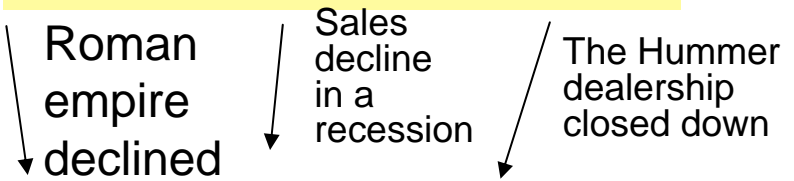
Similarity-based retrieval



Why did auto earnings decline?



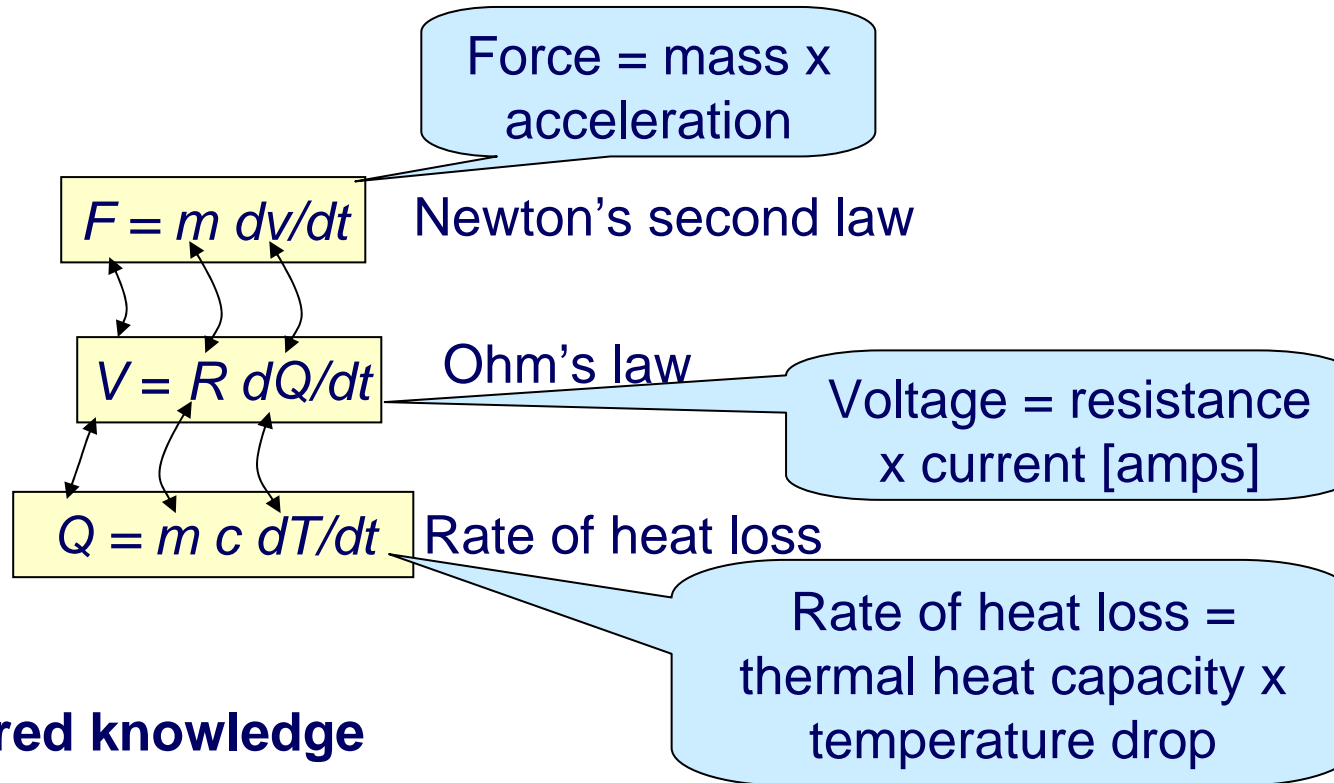
Structural mapping



Why did auto earnings decline?



Interdisciplinary Analogies



Transferred knowledge

Abstract concepts

“Flow variable” “Potential variable”

Abstract laws, relations

“Closed loop law” “Power” “Kinetic energy”

Conclusions

- What are the main characteristics of this approach in measuring creativity?
 - Ideational Fluency: Generate many responses to a prompt
 - Creativity: Fluency + hard to think of many responses
 - Scoring: #, # unusual, quality (e.g., plausible approaches)
- Where have the main advantages and disadvantages of this approach in a cross-national setting?
 - Advantages:
 - Accommodates correlational evidence; accommodates process model explanation (e.g., MAC/FAC; Case-based reasoning)
 - Lends itself to instruction/improvement
 - Disadvantages
 - Unknown what problems would be encountered cross culturally
 - Creativity's role in multilevel framework predicting innovation unknown

References

- Klenk, M. and Forbus, K. (2006). Analogical Model Formulation for AP Physics Problems. 20th International Workshop on Qualitative Reasoning, Hanover, NH.
- Forbus, K., Gentner, D., and Law, K. MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, 19, 141-205. 1994.
- Gentner, D., Structure-mapping: A theoretical framework for analogy, *Cognitive Science* 7(2), 1983.
- Frederiksen, N., & Ward, W.C. (1978). Measures for the study of creativity in scientific problem-solving. *Applied Psychological Measurement*, 2(1), 1-24.
- Bennett, R. E., & Rock, D.A., (1993). Generalizability, validity, and examinee perceptions of a computer-delivered formulating hypotheses test.



Adapting Creativity Tests for Cross-National Assessment

Patrick Kyllonen

Center for New Constructs (R&D)

Educational Testing Service

"CAN CREATIVITY BE MEASURED?"

28-29 May 2009

Brussels, MADOU auditorium

**Session 3 : Approaches to the
measurement of creativity at the
individual level (May 29, 1300—1430)**

Kyllonen, P. C. (2009). The tried, the true, and the transpiring in creativity assessment. In E. Villalba (organizer), *Can creativity be measured?..* OECD: Brussels, Belgium.

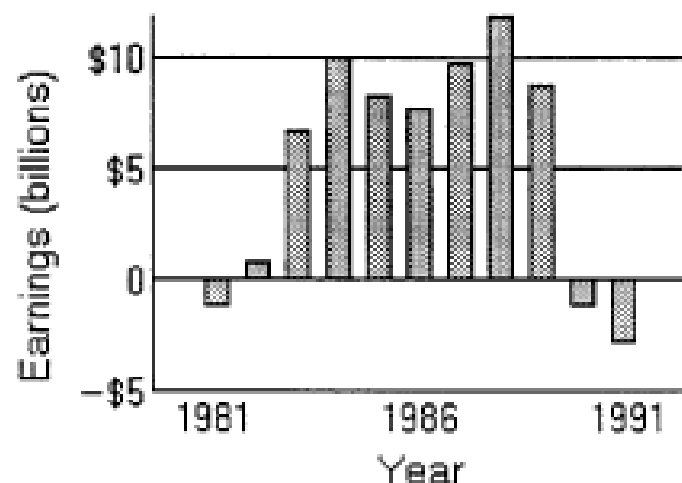
Guiding Questions

- How would different methods in the measurement of creativity be adapted in order to allow for cross-national comparisons?
- What are the steps that would be necessary to create a international survey?

More Available

SAMPLE PROBLEM 2

Combined Earnings for the Three Largest Automobile Manufacturers in Country X From 1981 to 1991



From 1983-1989, the earnings of the three largest automobile manufacturers in Country X never fell below 5 billion dollars a year. In 1990, the companies suffered a combined loss of 1.2 billion dollars, and in 1991 the combined loss was over 3 billion dollars.

Think of hypotheses that might explain the losses in 1990 and 1991 in Country X. Write each hypothesis as a separate answer of no more than **7 words**.

When finished reading directions click on the icon below

Dismiss Directions

Test



Time

Quit



Help

Answer

Confirm



Next