

Leverage Points for Improving Educational Assessment

Robert J. Mislevy, Linda S. Steinberg,
and Russell G. Almond

Educational Testing Service
February 25, 2000

Presented at the Technology Design Workshop sponsored by the U.S. Department of Education, held at Stanford Research Institute, Menlo Park, CA, February 25-26, 2000.

The work of the first author was supported in part by the Educational Research and Development Centers Program, PR/Award Number R305B60002, as administered by the Office of Educational Research and Improvement, U.S. Department of Education. The findings and opinions expressed in this report do not reflect the positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment, the Office of Educational Research and Improvement, or the U.S. Department of Education.

Some opportunities...

Cognitive/educational psychology

- » how people learn,
- » organize knowledge,
- » put knowledge to use.

Technology to...

- » create, present, and vivify “tasks”;
- » evoke, capture, parse, and store data;
- » evaluate, report, and use results.

A Challenge

- ✿ How the heck do you make sense of rich, complex data, for more ambitious inferences about students?

A Response

Design assessment from

generative principles ...

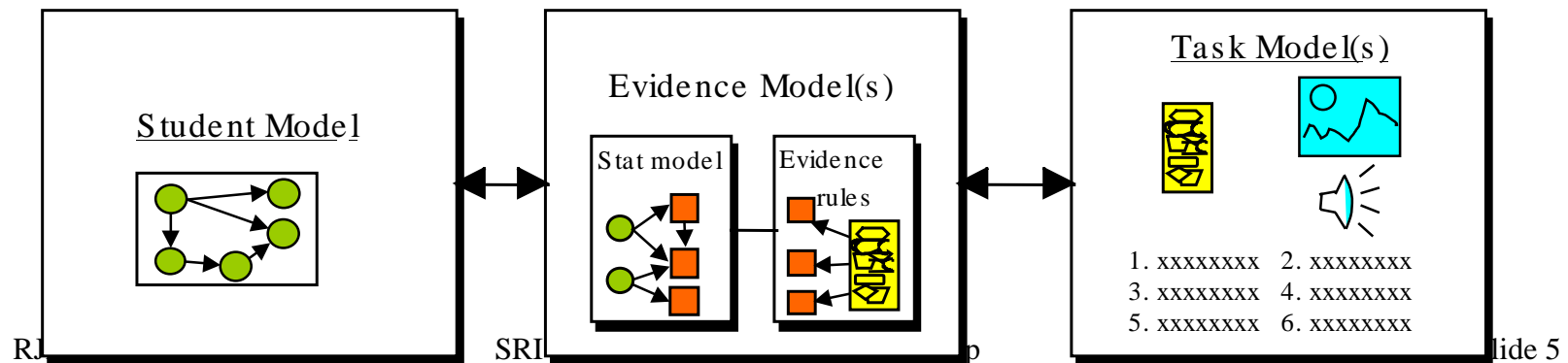
1. Psychology
2. Purpose
3. Evidentiary reasoning

Conceptual design LEADS

Tasks, statistics & technology FOLLOW

Principled Assessment Design

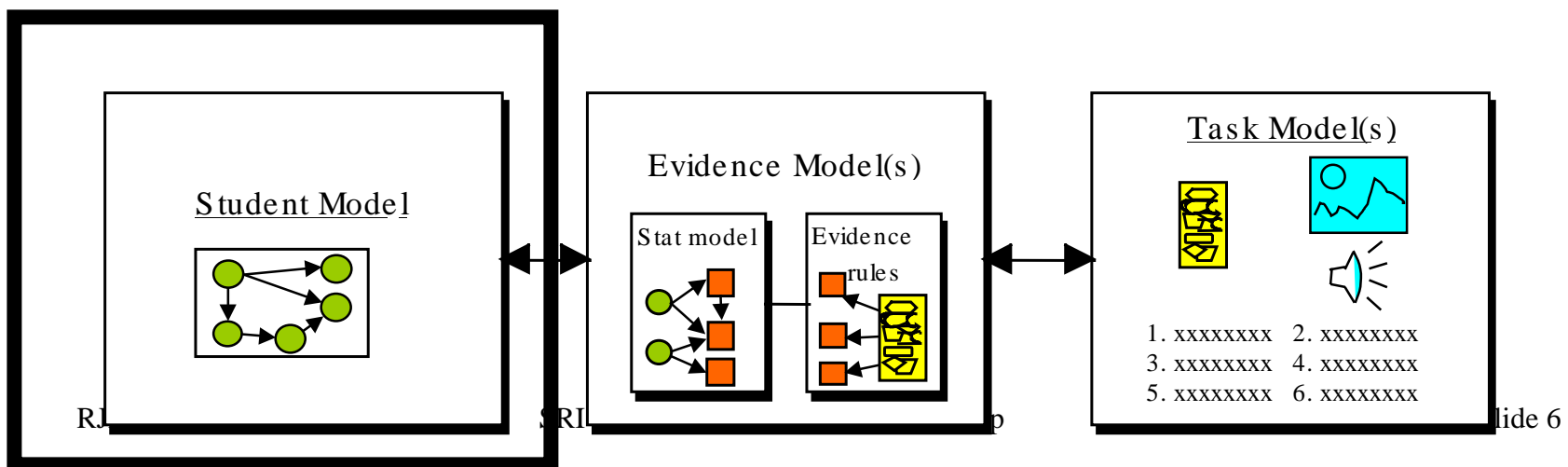
The three basic models



Evidence-centered assessment design

- ✿ What complex of knowledge, skills, or other attributes should be assessed, presumably because they are tied to explicit or implicit objectives of instruction or are otherwise valued by society?

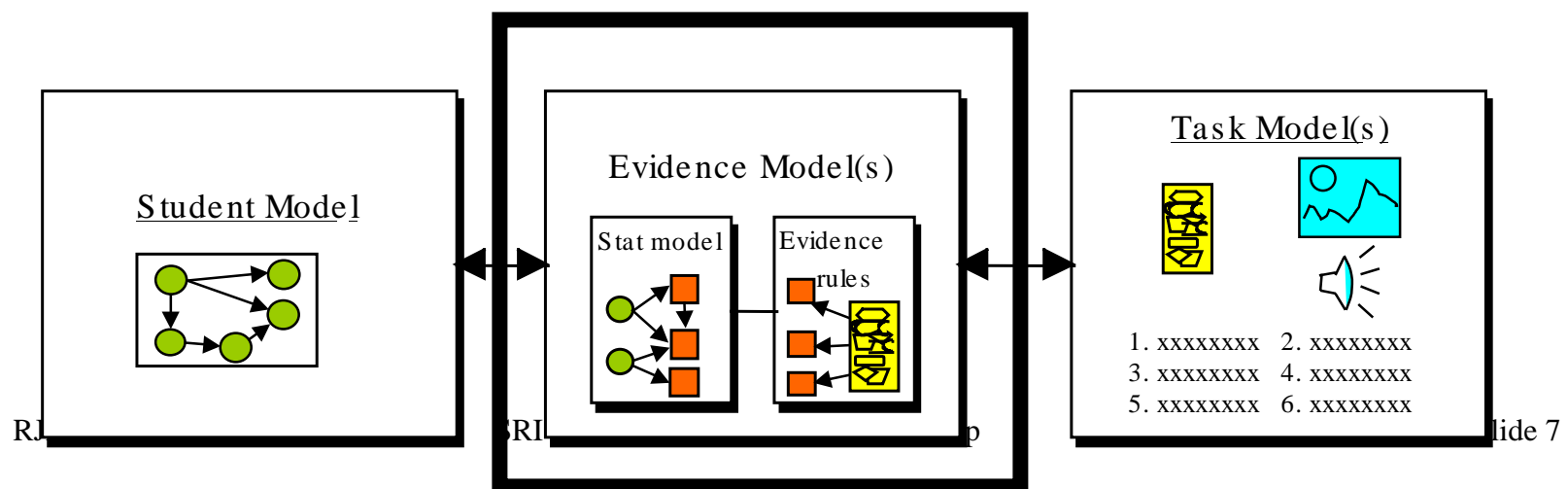
(Messick, 1992)



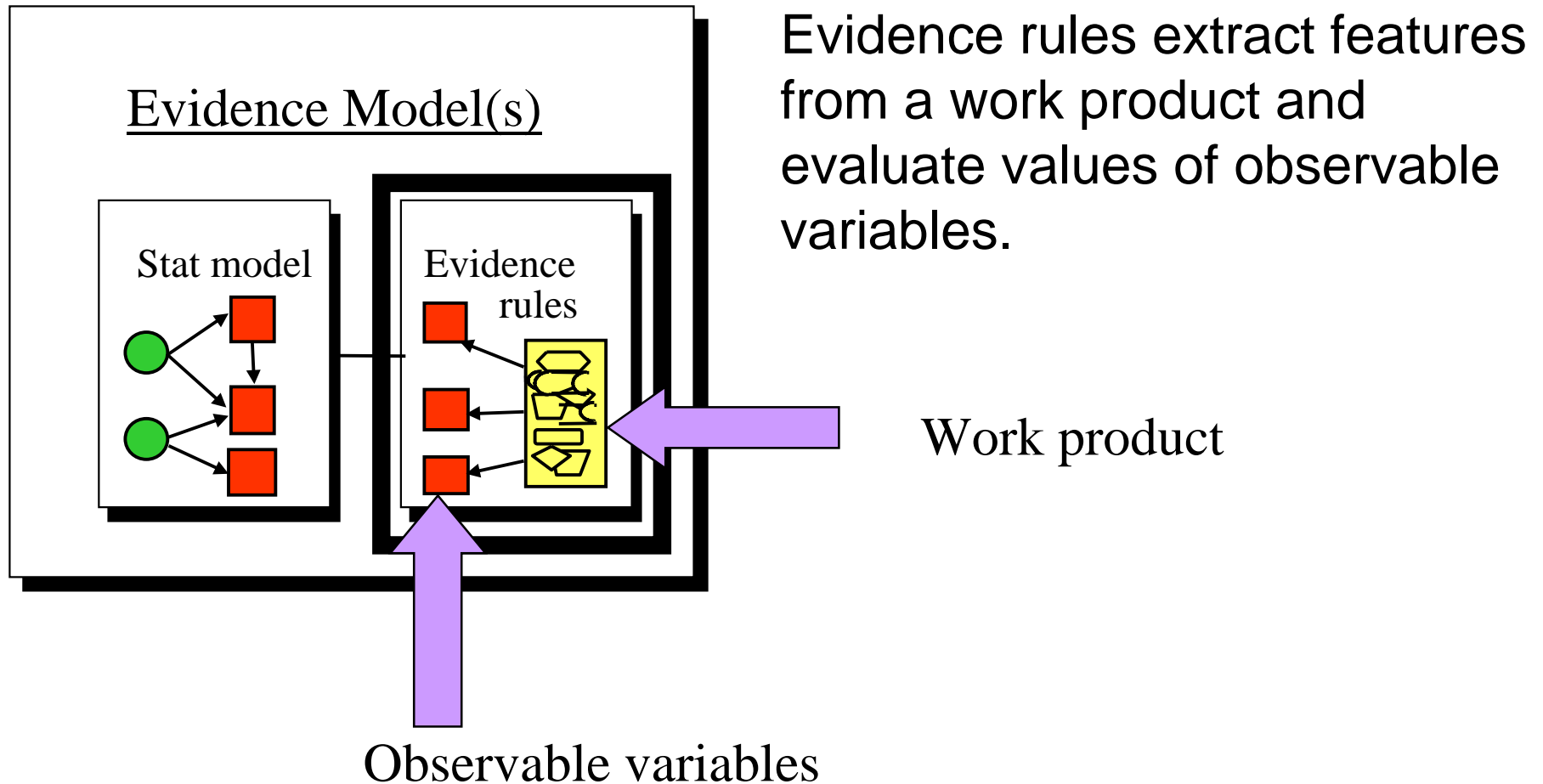
Evidence-centered assessment design

- ✿ What complex of knowledge, skills, or other attributes should be assessed, presumably because they are tied to explicit or implicit objectives of instruction or are otherwise valued by society?
- ✿ What behaviors or performances should reveal those constructs?

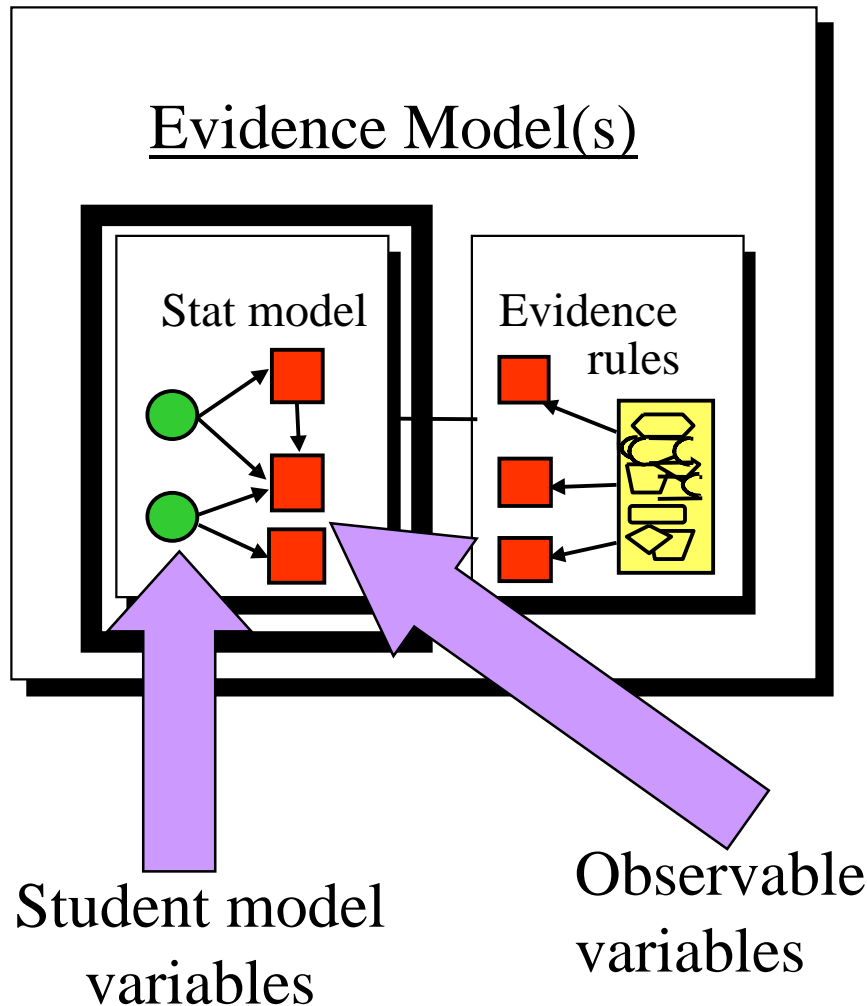
(Messick, 1992)



The Evidence Model(s)



The Evidence Model(s)

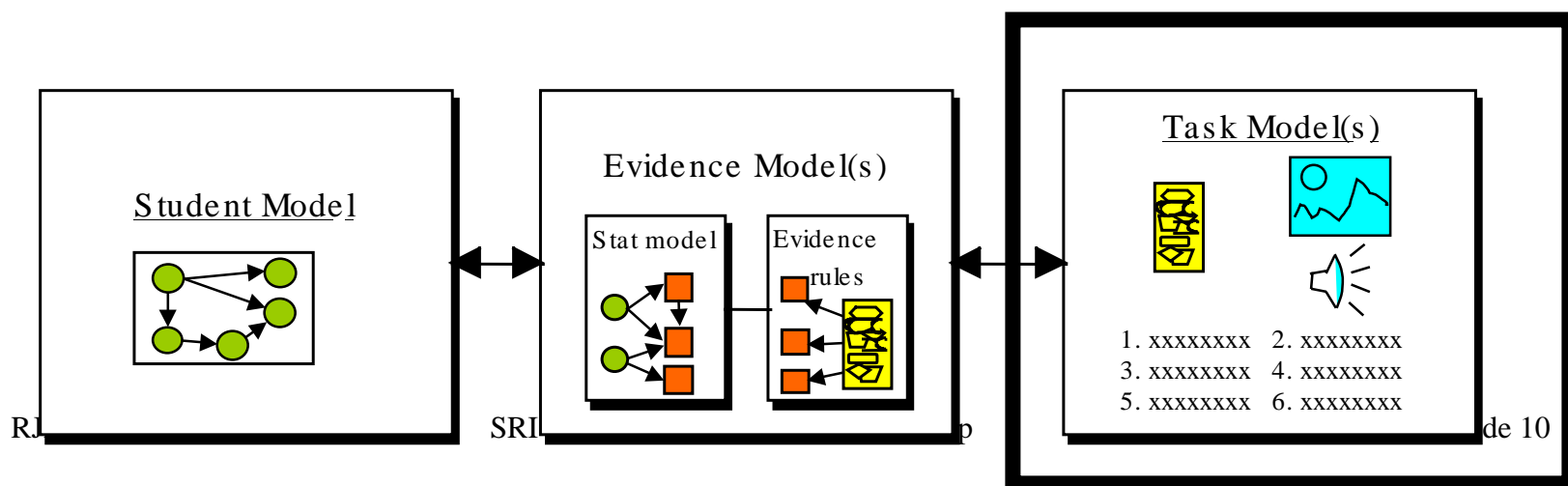


The statistical component expresses the how the observable variables depend, in probability, on student model variables.

Evidence-centered assessment design

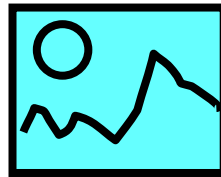
- ✿ What complex of knowledge, skills, or other attributes should be assessed, presumably because they are tied to explicit or implicit objectives of instruction or are otherwise valued by society?
- ✿ What behaviors or performances should reveal those constructs?
- ✿ What tasks or situations should elicit those behaviors?

(Messick, 1992)



The Task Model(s)

Task Model(s)

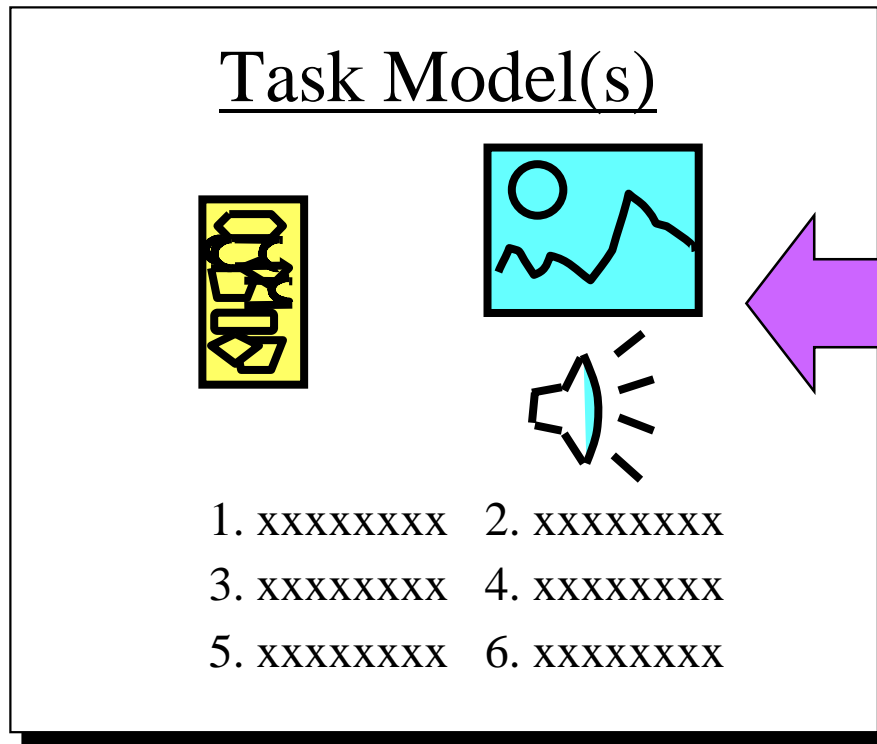


- | | |
|-------------|-------------|
| 1. xxxxxxxx | 2. xxxxxxxx |
| 3. xxxxxxxx | 4. xxxxxxxx |
| 5. xxxxxxxx | 6. xxxxxxxx |

Task-model variables describe features of tasks.

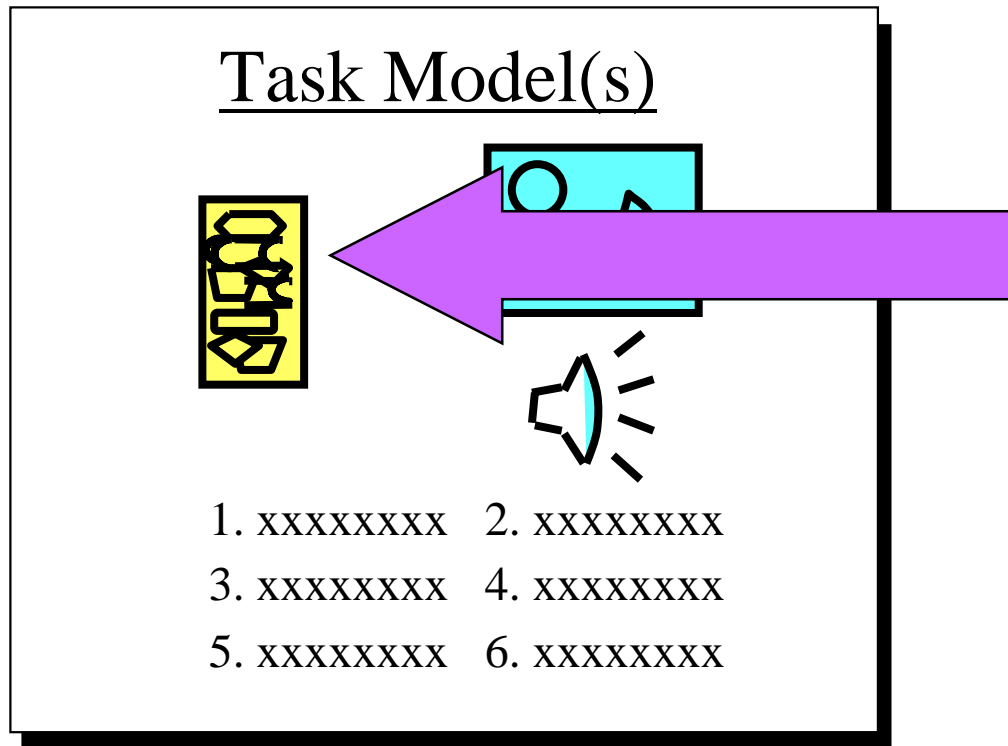
A task model provides a framework for describing and constructing the situations in which examinees act.

The Task Model(s)



Includes specifications for the stimulus material, conditions, and affordances-- the environment in which the student will say, do, or produce something.

The Task Model(s)



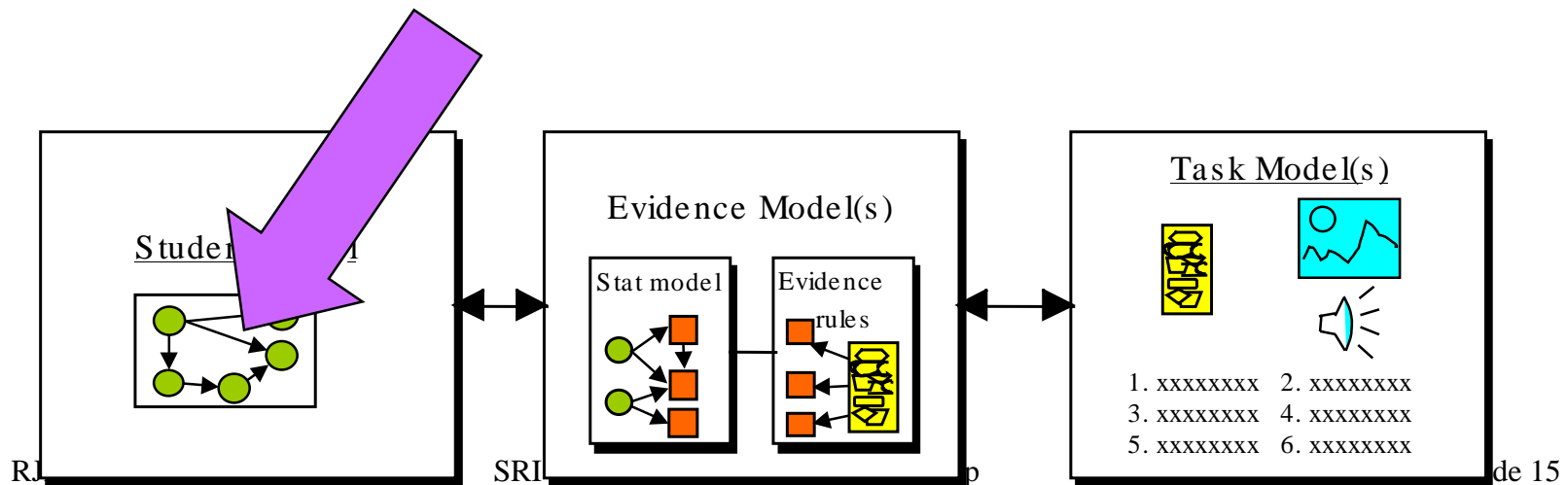
Includes specifications for the “work product”: the form in which what the student says, does, or produces will be captured.

Leverage Points...

- ✿ For cognitive/educational psychology
- ✿ For statistics
- ✿ For technology

Leverage Points for Cog Psych

- ✿ The character and substance of the student model.



Example a: GRE Verbal Reasoning

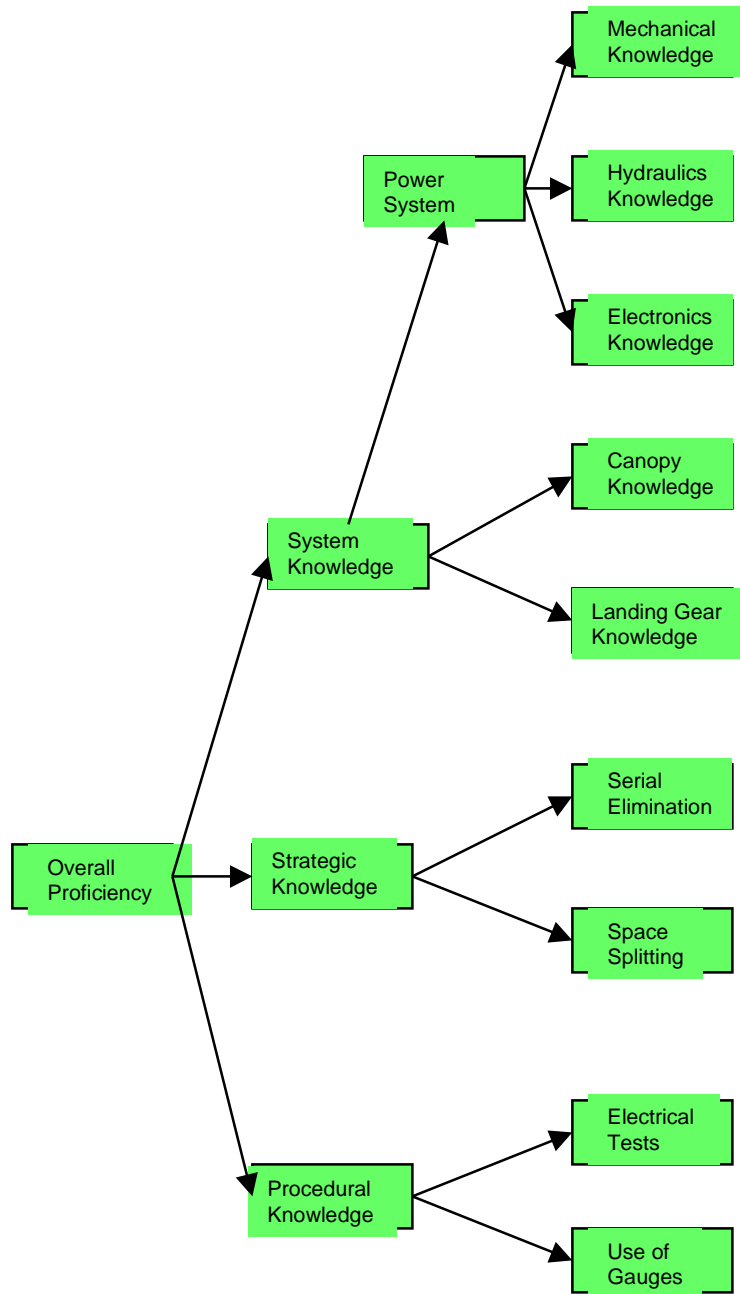
θ

The student model is just the IRT ability parameter θ —
the tendency to make correct responses in the mix of items presented in a GRE-V.

Example b: HYDRIVE

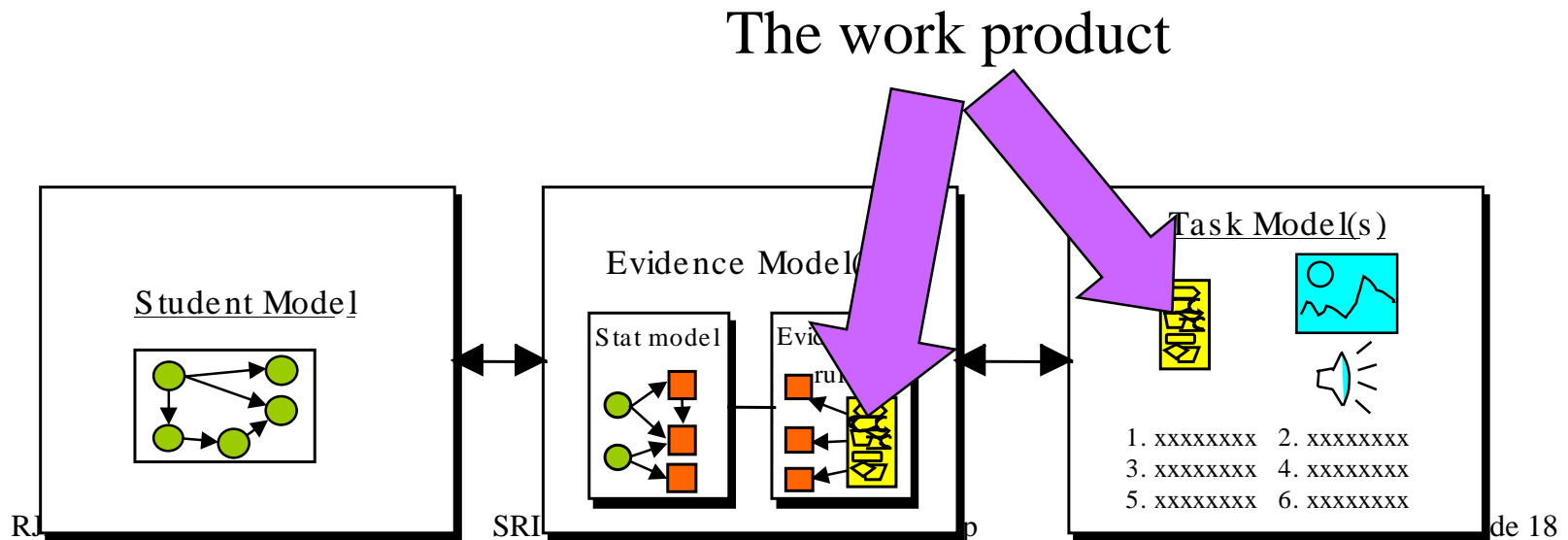
Student-model variables
in HYDRIVE

A Bayes net fragment.



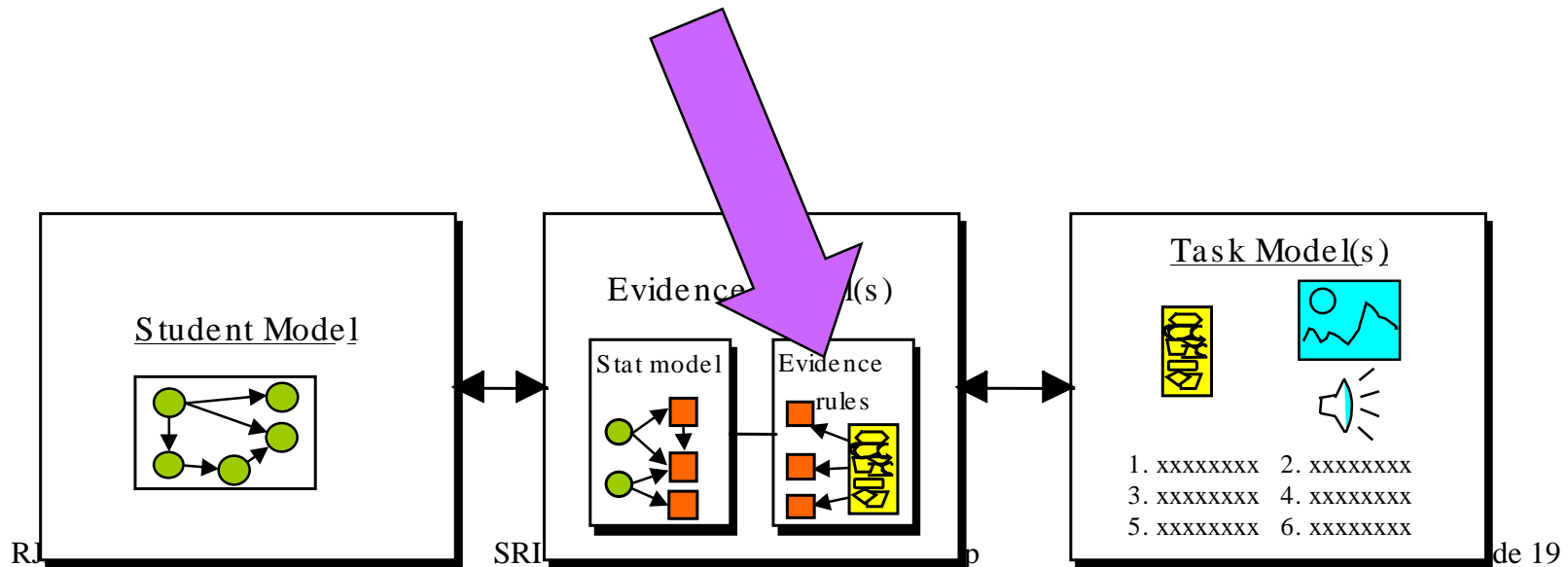
Leverage Points for Cog Psych

- ✿ The character and substance of the student model.
- ✿ What we can observe to give us evidence,



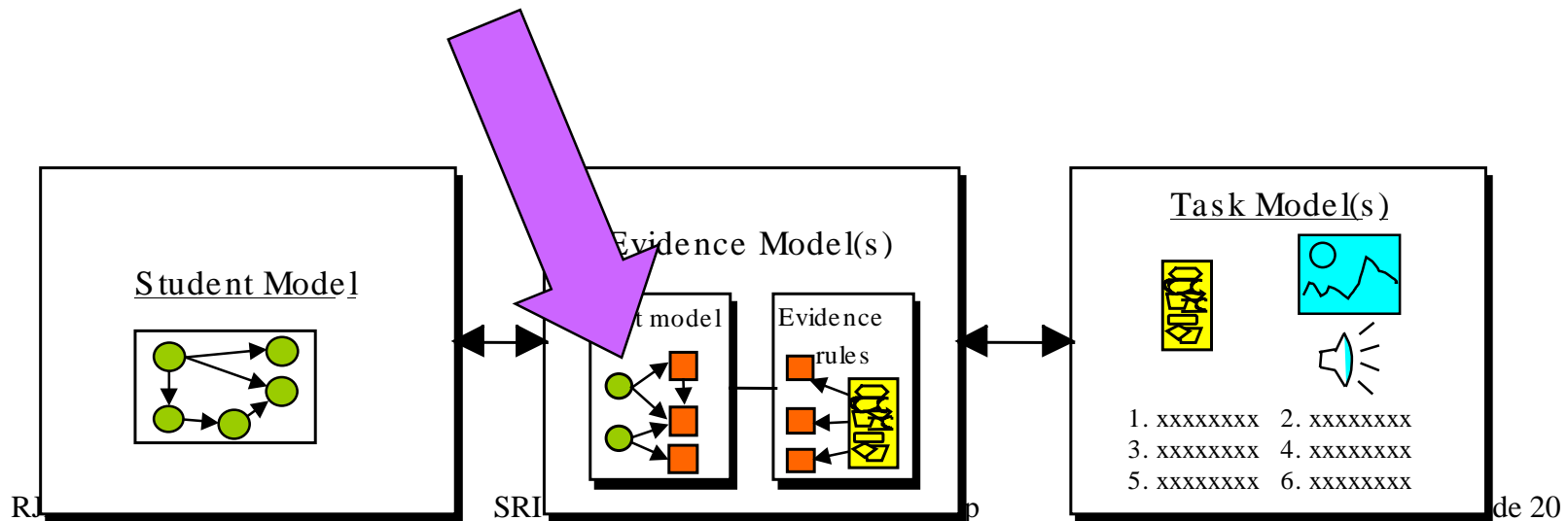
Leverage Points for Cog Psych

- ✿ The character and substance of the student model.
- ✿ What we can observe to give us evidence, and how to recognize and summarize its key features.



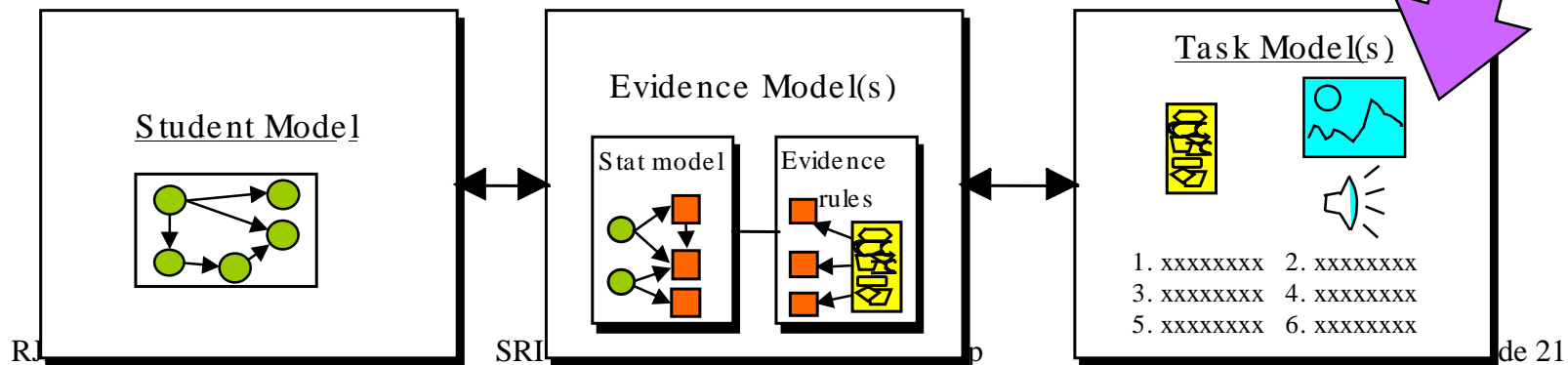
Leverage Points for Cog Psych

- ✿ The character and substance of the student model.
- ✿ What we can observe to give us evidence, and how to recognize and summarize its key features.
- ✿ Modeling which aspects of performance depend on which aspects of knowledge, in what ways.



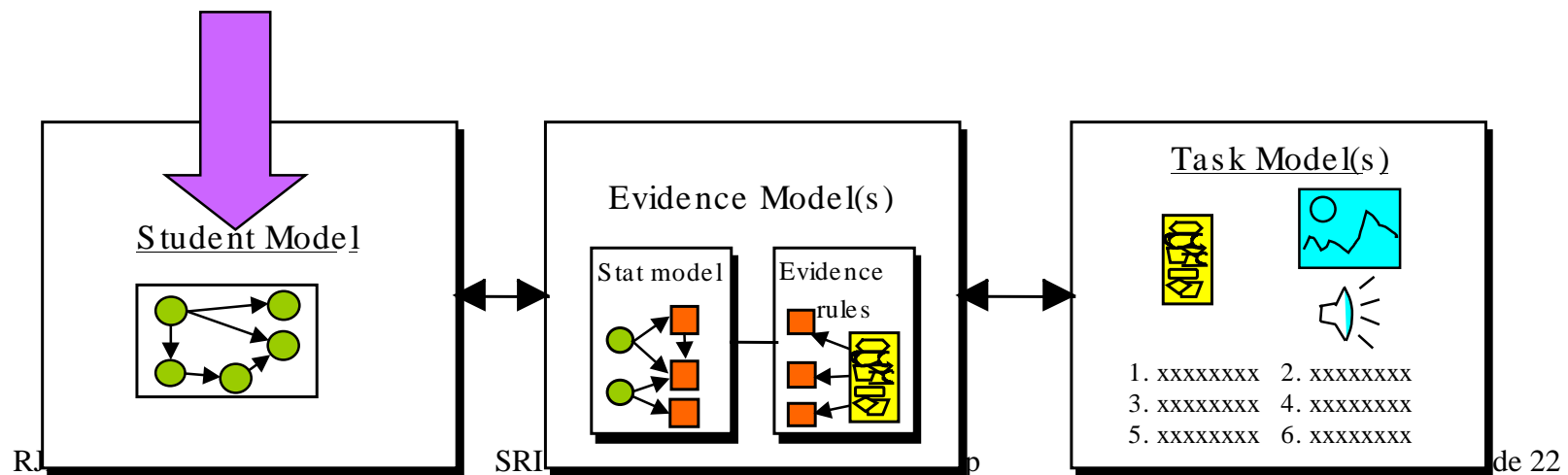
Leverage Points for Cog Psych

- ✿ The character and substance of the student model.
- ✿ What we can observe to give us evidence, and how to recognize and summarize its key features.
- ✿ Modeling how which aspects of performance depend on which aspects of knowledge , in what ways.
- ✿ Effective ways to elicit the kinds of behavior we need to see.



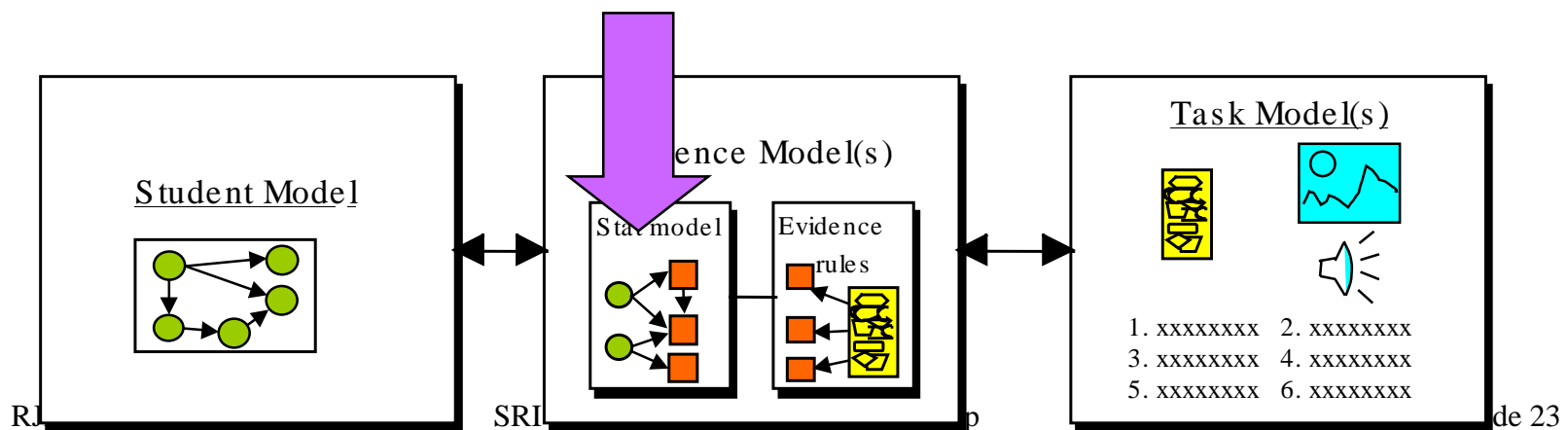
Leverage Points for Statistics

- * Managing uncertainty with respect to the student model.
 - * Bayes nets (generalize beyond familiar test theory models--eg, VanLehn)
 - * Modular construction of models
 - * Monte Carlo estimation
 - * Knowledge-based model construction wrt the student model.

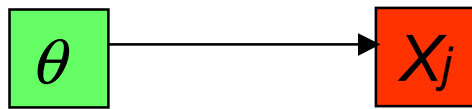


Leverage Points for Statistics

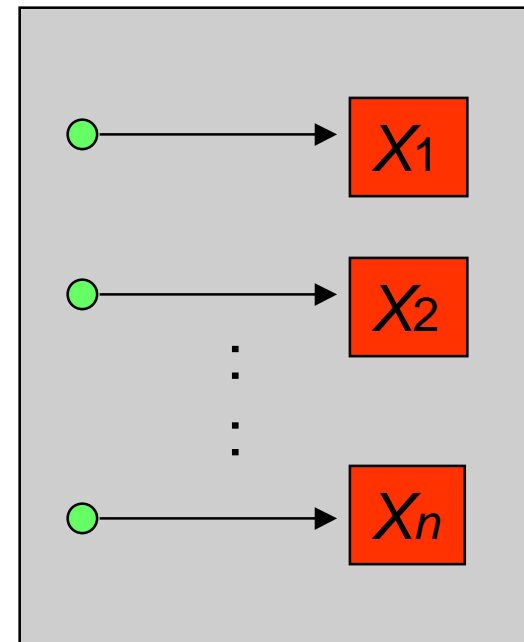
- * Managing the stochastic relationship between observations in particular tasks and the persistent unobservable student model variables.
 - * Bayes nets
 - * Modular construction of models (incl psychometric building blocks)
 - * Monte Carlo approximation
 - * Knowledge-based model construction--docking with the student model.



Example a, continued: GRE-V



Sample Bayes net --
Student model fragment
docked with an
Evidence Model fragment
(IRT model & parameters for this item)

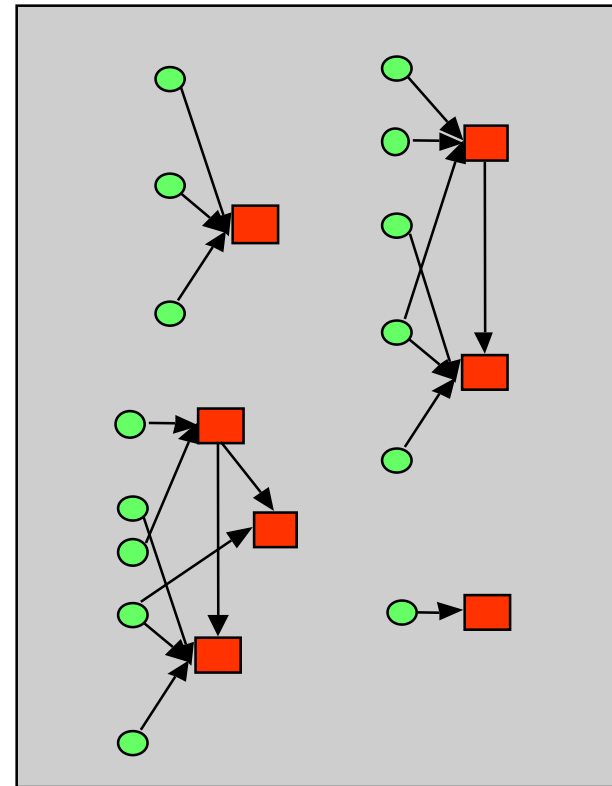


Library of
Evidence Model
Bayes net fragments

Example b, continued: HYDRIVE



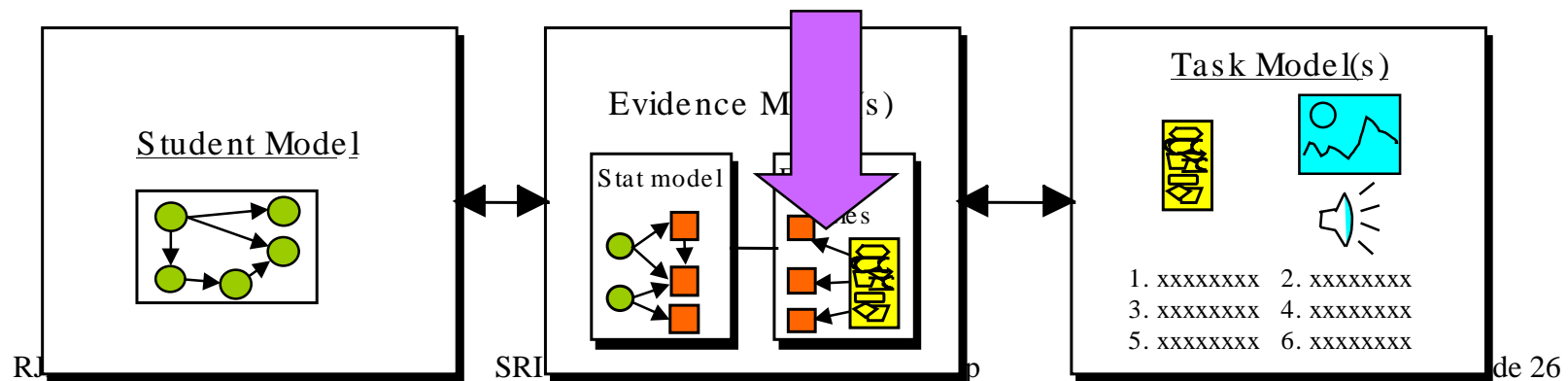
Sample Bayes net fragment



Library of fragments

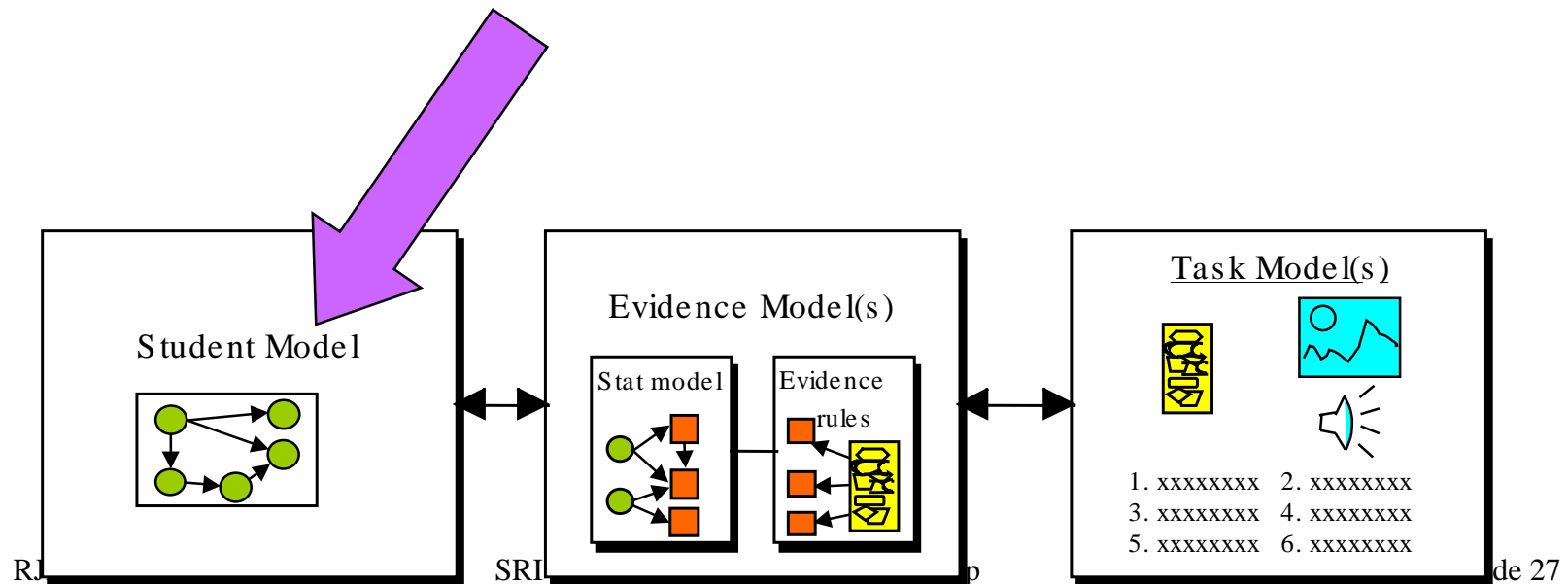
Leverage Points for Statistics

- * Extracting features and determining values of observable variables .
 - * Bayes nets (also neural networks, rule-based logic)
 - * Modeling human raters for training, quality control, efficiency



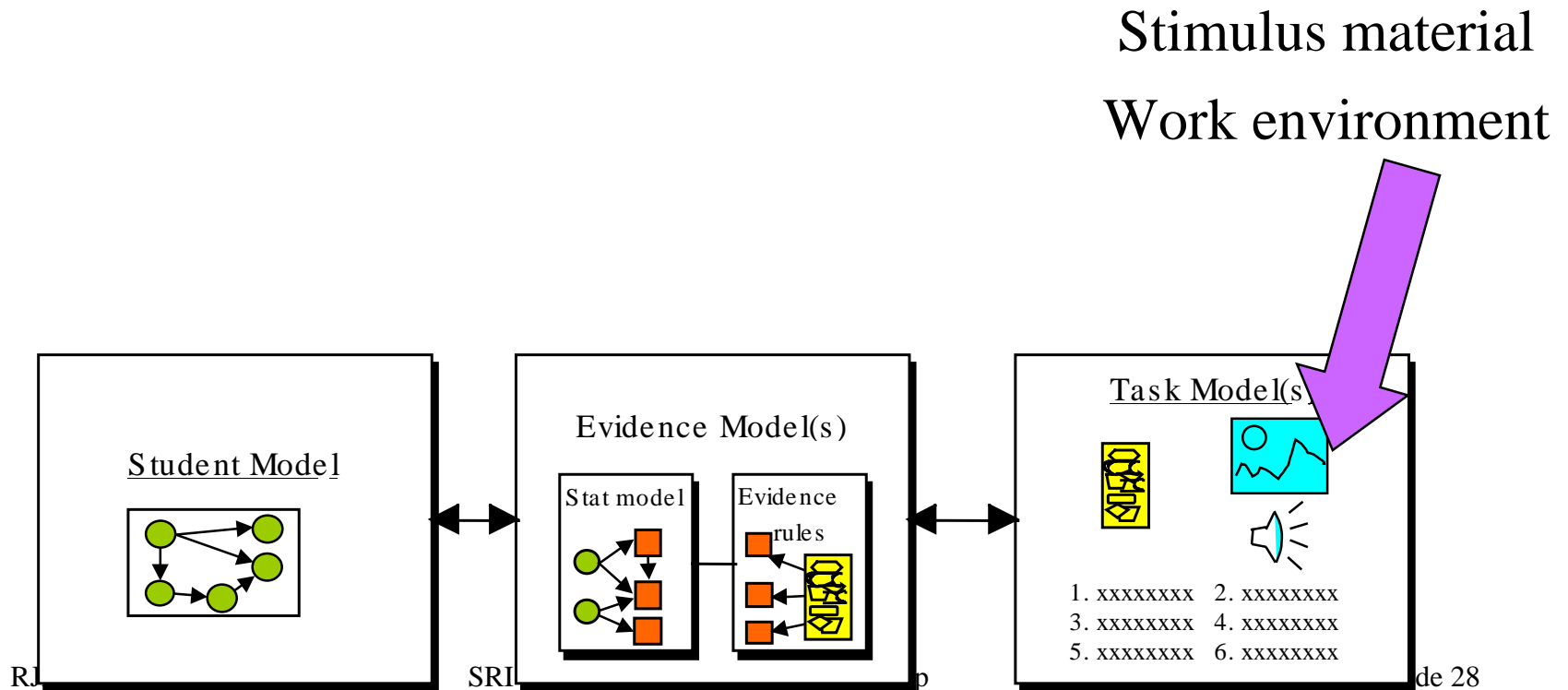
Leverage Points for Technology

- ✿ Dynamic assembly of the student model.



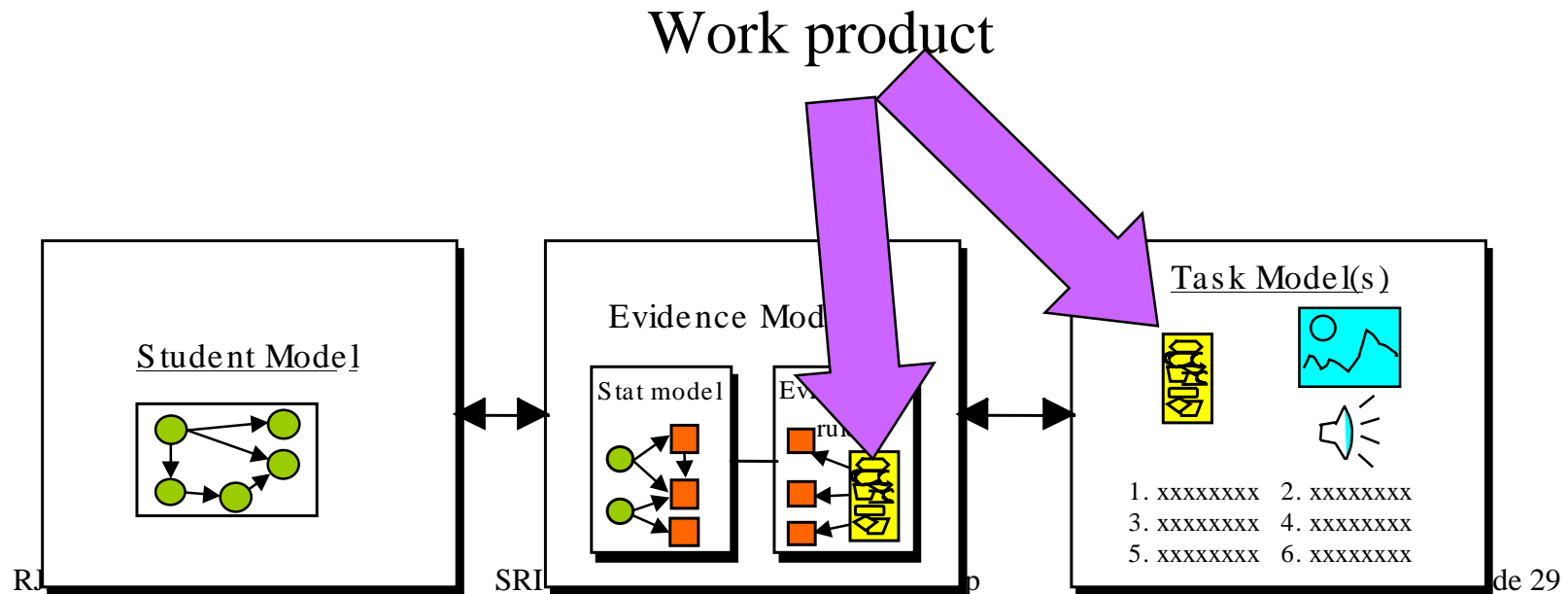
Leverage Points for Technology

- ✿ Dynamic assembly of the student model.
- ✿ Complex and realistic tasks that can produce direct evidence about knowledge used for production and interaction.



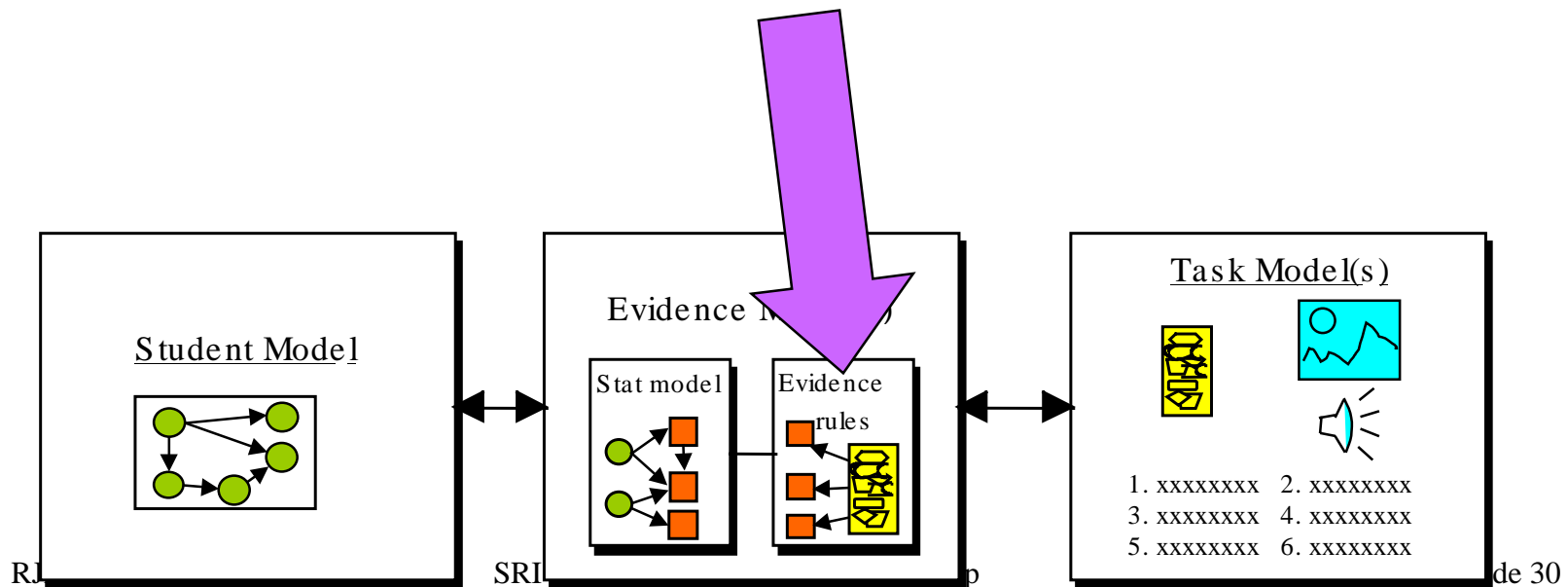
Leverage Points for Technology

- ✿ Dynamic assembly of the student model.
- ✿ Complex and realistic tasks that can produce direct evidence about knowledge used for production and interaction.



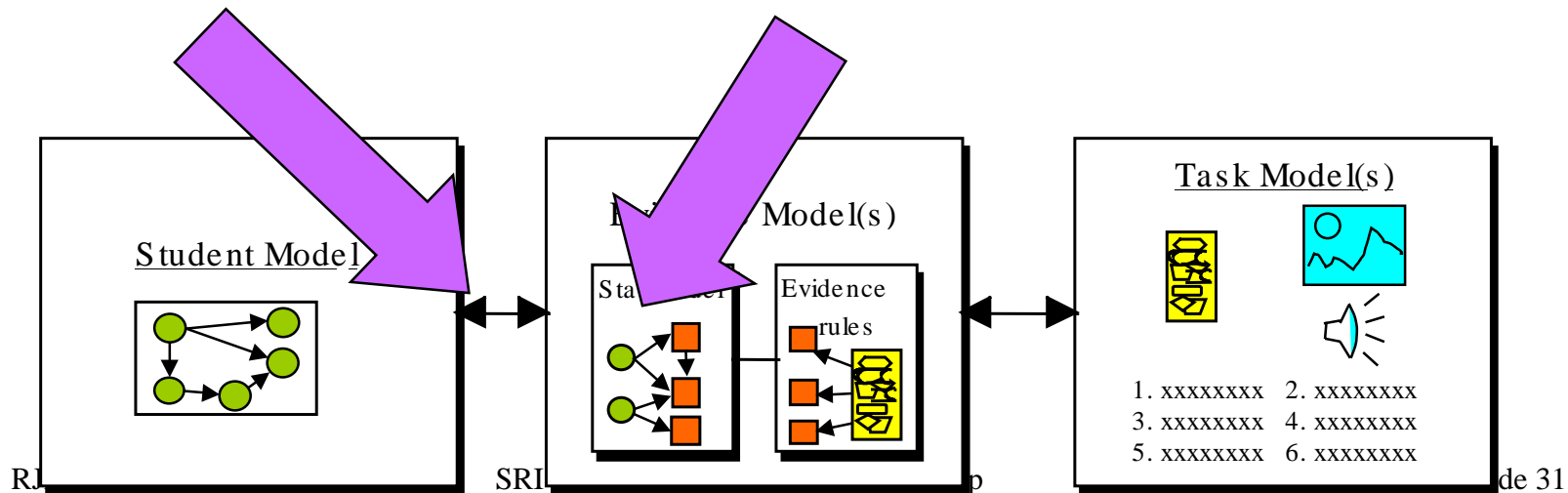
Leverage Points for Technology

- ✿ Dynamic assembly of the student model.
- ✿ Complex and realistic tasks that can produce direct evidence about knowledge used for production and interaction.
- ✿ Automated extraction and evaluation of key features of complex work.



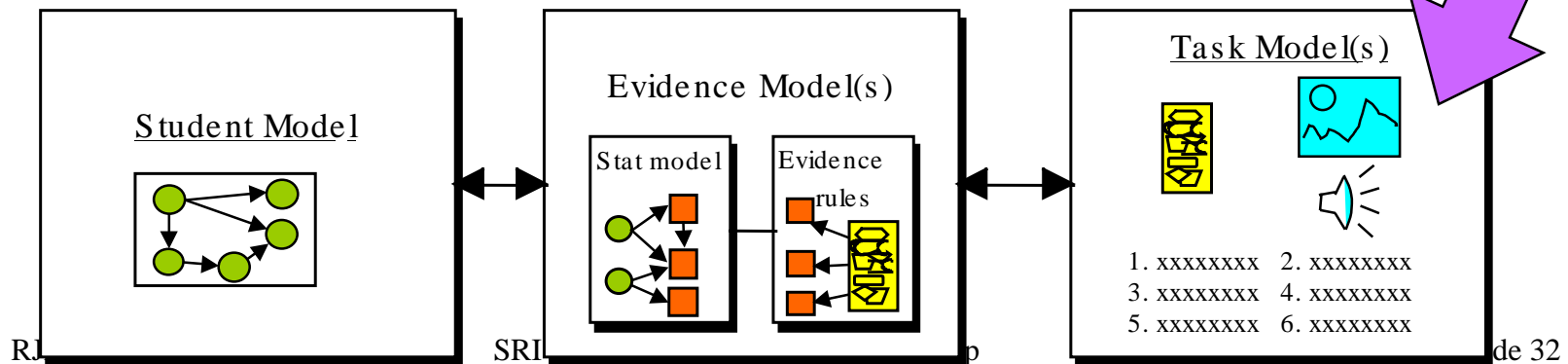
Leverage Points for Technology

- * Dynamic assembly of the student model.
- * Complex and realistic tasks that can produce direct evidence about knowledge used for production and interaction.
- * Automated extraction and evaluation of key features of complex work.
- * Construction and calculation to guide acquisition of, and manage of uncertainty about, our knowledge about the student.



Leverage Points for Technology

- ✿ Dynamic assembly of the student model.
- ✿ Complex and realistic tasks that can produce direct evidence about knowledge used for production and interaction.
- ✿ Automated extraction and evaluation of key features of complex work.
- ✿ Construction and calculation to guide acquisition of, and manage and uncertainty about, knowledge about the student.
- ✿ Automated/assisted task construction, presentation, management



The Cloud behind the Silver Lining

- ✿ These developments will have the *most* impact when assessments are built for well-defined purposes, and connected with a conception of knowledge in the targeted domain.
- ✿ They will have *much less* impact for ‘drop-in-from-the-sky’ large-scale assessments like NAEP.