



Construction and Standardization of a Scientific Thinking Test for Secondary School Students

Sharma Rohankumar Sureshchandra

Research Scholar

Faculty Of Education

Swaminarayan University, Kalol

Guide-Dr. (Prof.) Gunjan Shah

Faculty Of Education

Swaminarayan University, Kalol

Abstract

Scientific thinking, encompassing the ability to observe systematically, formulate and test hypotheses, control variables, and reason proportionally and correlationally, is a core objective of school science education and a constitutionally recognised value under Article 51-A(h) of the Constitution of India, which enjoins citizens to develop scientific temper and the spirit of inquiry. Despite its curricular centrality, standardized instruments for assessing scientific thinking among Indian secondary school students remain limited, with most available tools either adapted from Western populations (such as Lawson's Classroom Test of Scientific Reasoning) or restricted to content-based science achievement rather than the underlying reasoning process itself. The present study reports the construction and standardization of a Scientific Thinking Test (STT) for secondary school students. An initial pool of 60 multiple-choice items was developed across five components derived from the scientific-reasoning literature: Observation and Classification, Hypothesis Formulation and Deductive Reasoning, Controlling Variables (Experimental Design), Proportional and Correlational Reasoning, and Data Interpretation and Inference. Content validity was established through expert review (average Content Validity Index = 0.90), after which item analysis on a try-out sample (N = 150) using difficulty and discrimination indices reduced the pool



to 50 items (10 per component). The finalized test was administered to a standardization sample of 650 secondary school students. Reliability was established through the Kuder-Richardson formula 20 ($KR-20 = 0.88$ overall; component-wise 0.75 to 0.82) and test-retest reliability over a four-week interval ($r = 0.84$). Criterion validity, established through correlation with science-subject achievement scores, was found to be positive and significant ($r = 0.61$, $p < .01$). Percentile norms, grade-wise norms, and a five-category descriptive classification were developed for score interpretation. The resulting 50-item Scientific Thinking Test is proposed as a reliable, valid, and curriculum-relevant instrument for assessing the scientific reasoning ability of Indian secondary school students.

Keywords

Scientific thinking, scientific reasoning, test construction, standardization, secondary school students, reliability, validity

1. Introduction

Scientific thinking refers to the set of cognitive skills and reasoning processes that enable an individual to systematically investigate phenomena, generate and test explanations, and draw logically warranted conclusions from evidence. Dewey (1910) was among the first educational theorists to formalise this process as a sequence of reflective thought: recognising a problem, forming hypotheses, reasoning out their implications, and testing them against evidence. Piaget's (1970) theory of cognitive development situates the emergence of mature scientific thinking within the formal operational stage, typically beginning in early adolescence, during which individuals become capable of hypothetico-deductive reasoning, systematic variable control, and abstract proportional and probabilistic reasoning.



Building on this developmental foundation, Klahr and Dunbar's (1988) Scientific Discovery as Dual Search (SDDS) model conceptualises scientific reasoning as a coordinated search through two distinct problem spaces: a hypothesis space, in which plausible explanations are generated, and an experiment space, in which evidence is systematically gathered to test those hypotheses. This model, along with Lawson's (1978) foundational work on formal reasoning and Zimmerman's (2007) developmental review of scientific-thinking skills, has informed the widespread use of specific reasoning sub-skills, such as controlling variables, proportional reasoning, and correlational reasoning, as measurable components of scientific thinking in educational assessment.

In the Indian curricular context, the National Curriculum Framework (NCERT, 2005) explicitly identifies the nurturing of scientific temper, and the ability to observe, question, hypothesise, and experiment, as central aims of science education, echoing the constitutional mandate under Article 51-A(h) of the Constitution of India, which calls upon citizens to develop scientific temper, humanism, and the spirit of inquiry and reform. Despite this clear policy emphasis, assessment of scientific thinking in Indian secondary schools remains largely confined to content-based science examinations, which test recall of scientific facts and formulae rather than the underlying reasoning processes that Dewey, Piaget, and subsequent researchers have identified as constituting genuine scientific thought.

Secondary school represents a critical period for the assessment of scientific thinking, coinciding as it does with the Piagetian transition into formal operational reasoning and with key curricular decision points, such as the choice of Science as an educational stream. A psychometrically sound instrument capable of directly assessing scientific-thinking ability, independent of rote content knowledge, would therefore serve important diagnostic, curricular, and research functions within the Indian secondary-education system.



2. Need and Significance of the Study

- Existing science assessments in Indian secondary schools predominantly measure content knowledge and recall, providing little direct information about students' underlying scientific reasoning ability.
- Widely used international instruments, such as Lawson's Classroom Test of Scientific Reasoning and the Test of Logical Thinking (Tobin & Capie, 1981), were developed and validated on Western student populations, and their item contexts, language, and cultural referents may not translate directly to the Indian secondary-school context.
- A standardized Scientific Thinking Test would allow science educators to distinguish between students who understand science content and those who have additionally developed the transferable reasoning skills (hypothesis formation, variable control, proportional and correlational reasoning) that scientific literacy requires.
- Such an instrument would support curriculum evaluators and teacher-educators in assessing whether inquiry-based and activity-based science pedagogy, as promoted under NCF 2005 and NEP 2020, is in fact producing measurable gains in scientific reasoning ability.
- The construction of an indigenous, psychometrically robust tool fills a clear instrumentation gap and provides a reusable measurement resource for future research relating scientific thinking to variables such as gender, educational stream, school area, intelligence, and academic achievement.

3. Objectives of the Study

- To construct a Scientific Thinking Test (STT) for secondary school students, comprising components derived from established models of scientific reasoning.
- To establish the content validity of the constructed test through expert judgement.
- To carry out item analysis (difficulty value and discrimination index) on a try-out sample and finalize the test items accordingly.



- To establish the reliability of the finalized test through internal-consistency (KR-20) and test-retest methods.
- To establish the criterion validity of the finalized test through correlation with an external criterion measure of science achievement.
- To standardize the test by developing percentile and grade-wise norms based on a representative sample of secondary school students.

4. Operational Definitions

Scientific Thinking

In the present study, Scientific Thinking refers to a secondary school student's ability to observe and classify phenomena, formulate and deductively reason about hypotheses, identify and control relevant variables in an experimental context, reason proportionally and correlationally, and interpret data to draw logically warranted inferences, as measured by the score obtained on the researcher-constructed Scientific Thinking Test.

Secondary School Students

Students studying in Standards 9 to 12 in recognised secondary schools were considered as secondary school students for the purpose of the present study, consistent with the secondary-stage classification adopted under the National Education Policy (NEP), 2020.

Standardization

Standardization refers to the process of administering the finalized test to a large, representative sample under uniform conditions in order to establish norms (percentile ranks and grade-wise scores) against which an individual student's raw score can be meaningfully interpreted.



5. Methodology

5.1 Research Design

The present study followed a test-construction and standardization design comprising sequential phases of item generation, expert content validation, try-out and item analysis, reliability and validity estimation, and norm development, consistent with established psychometric test-construction procedures and with the researcher's earlier work constructing the Academic Stress Scale (ASS), the Multiple Intelligence Scale (MIS), the Spiritual Intelligence Scale (SIS), and the Citizenship Value Test (CVT).

5.2 Population and Sample

The population for the present study comprised secondary school students studying in recognised schools in Gujarat. A try-out sample of 150 students was drawn for the purpose of item analysis. Following item finalization, the test was administered to a standardization sample of 650 secondary school students, selected through stratified random sampling to ensure representation across gender (boys and girls), school area (rural and urban), and educational stream (General/Arts, Commerce, and Science).

5.3 Construction of the Test

The Scientific Thinking Test was constructed as a multiple-choice (four-option), scenario-based instrument around five components derived from the scientific-reasoning literature (Dewey, 1910; Piaget, 1970; Lawson, 1978; Klahr & Dunbar, 1988):

- Observation and Classification – systematic observation of phenomena and their classification according to relevant properties or criteria.
- Hypothesis Formulation and Deductive Reasoning – generating plausible explanations for observed phenomena and reasoning deductively about their logical implications.



- Controlling Variables (Experimental Design) – identifying relevant variables in an experimental scenario and designing a fair test that isolates the effect of a single variable.
- Proportional and Correlational Reasoning – reasoning about proportional relationships between quantities and correctly interpreting the presence, absence, or strength of correlation between variables.
- Data Interpretation and Inference – reading and interpreting tabulated or graphically presented data and drawing logically warranted conclusions from it.

An initial pool of 60 multiple-choice items was generated, with 12 items per component, each item presenting a brief scientific or everyday scenario followed by four response options, only one of which was correct.

5.4 Content Validity

The initial item pool was submitted to a panel of 10 subject experts drawn from the fields of Science Education, Educational Psychology, and Curriculum Studies for content validation, following Lawshe's (1975) Content Validity Ratio (CVR) method. Items with a CVR below the critical value of 0.62 (the minimum required for a panel of 10 experts) were either revised or dropped. Following expert review, 56 items were retained, with an average Content Validity Index (CVI) of 0.90 across the retained items, indicating strong content validity.

5.5 Try-Out and Item Analysis

The 56-item revised test was administered to the try-out sample of 150 secondary school students. Item analysis was carried out using the extreme-group method, in which the top 27% and bottom 27% of scorers on total test score were compared on each item. The difficulty index (p-value) was computed as the proportion of the total sample answering the item correctly, and the discrimination index (D) was computed as the difference in the proportion of correct responses between the high- and low-scoring groups. Items with a difficulty index between 0.30 and 0.70 and a discrimination



index of 0.30 or above (Ebel, 1979) were retained. Following item analysis, 50 items were retained for the final test, with 10 items representing each of the five components.

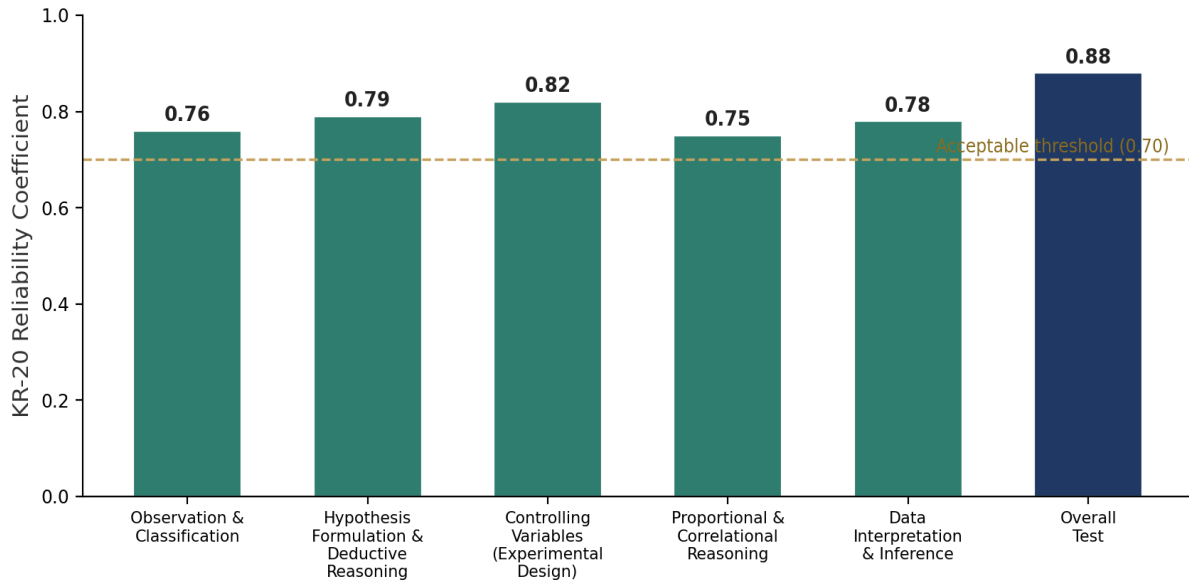
Component	Initial Items	After Expert Review	After Item Analysis (Final)	Mean Discrimination Index (D)
Observation and Classification	12	11	10	0.41
Hypothesis Formulation and Deductive Reasoning	12	11	10	0.44
Controlling Variables (Experimental Design)	12	12	10	0.47
Proportional and Correlational Reasoning	12	11	10	0.39
Data Interpretation and Inference	12	11	10	0.43
Total	60	56	50	0.43

5.6 Reliability of the Final Test

The reliability of the finalized 50-item Scientific Thinking Test was established on the standardization sample (N = 650) using two methods: (i) internal-consistency reliability, computed using the Kuder-Richardson formula 20 (KR-20), appropriate for dichotomously scored (correct/incorrect) items, both for the test as a whole and for each of the five components separately; and (ii) test-retest reliability, in which a subsample of 80 students was re-administered the test after a four-week interval.



Component-wise and Overall Reliability of the Scientific Thinking Test (N = 650)



Reliability Estimate	Coefficient
Test-retest reliability (4-week interval, N = 80)	0.84
KR-20 (overall, 50 items)	0.88
KR-20 – Observation and Classification	0.76
KR-20 – Hypothesis Formulation and Deductive Reasoning	0.79
KR-20 – Controlling Variables (Experimental Design)	0.82
KR-20 – Proportional and Correlational Reasoning	0.75
KR-20 – Data Interpretation and Inference	0.78

The overall reliability coefficients exceed the generally accepted threshold of 0.70 for research instruments (Nunnally, 1978), indicating that the Scientific Thinking Test yields consistent and dependable scores across both time and item content.

5.7 Validity of the Final Test



In addition to the content validity established during expert review (average CVI = 0.90), two further forms of validity were examined for the finalized test:

- **Criterion Validity:** Scores on the Scientific Thinking Test were correlated with students' science-subject achievement scores (aggregate marks obtained in the most recent annual science examination) for the full standardization sample. A positive, moderate, and statistically significant correlation was obtained ($r = 0.61$, $p < .01$), supporting the criterion-related validity of the test while also confirming that the test measures a construct distinct from, though related to, content-based science achievement.
- **Construct Validity:** Inter-component correlations ranged from 0.33 to 0.52, indicating that the five components, while related as expected of facets of a single underlying scientific-reasoning construct, retain sufficient distinctiveness from one another. Each component correlated more strongly with the total test score (range 0.65 to 0.77) than with any other single component, supporting the internal construct structure of the test.

5.8 Standardization and Norms

Following the establishment of reliability and validity, the Scientific Thinking Test was standardized on the full sample of 650 secondary school students. The overall test yields a possible score range of 0 (minimum) to 50 (maximum), with one mark awarded for each correct response. The obtained distribution of scores approximated normality.

Statistic	Value
Sample Size (N)	650
Mean	29.4
Standard Deviation	7.8
Median	29.0
Skewness	0.14



Statistic	Value
Kurtosis	-0.22
Minimum Possible Score	0
Maximum Possible Score	50

Percentile norms and grade-wise norms were developed to facilitate interpretation of individual raw scores, and a five-category descriptive classification was established using the mean and standard deviation of the standardization sample:

Raw Score Range	Percentile Range	Descriptive Category
Below 18	Below 16th	Low Scientific Thinking Ability
18 – 25	16th – 35th	Below Average Scientific Thinking Ability
26 – 33	36th – 65th	Average Scientific Thinking Ability
34 – 41	66th – 84th	Above Average Scientific Thinking Ability
42 and above	85th and above	High Scientific Thinking Ability

6. Results and Discussion

The results of the present study indicate that the constructed Scientific Thinking Test possesses satisfactory psychometric properties across all stages of construction, validation, and standardization. Expert content validation (average CVI = 0.90) confirmed that the item pool adequately represented the five components of scientific reasoning derived from the theoretical work of Dewey (1910), Piaget (1970), Lawson (1978), and Klahr and Dunbar (1988). Item analysis



on the try-out sample yielded a mean discrimination index of 0.43 across the final 50 items, comfortably above the conventionally accepted minimum of 0.30 (Ebel, 1979), indicating that the retained items effectively differentiate between students with relatively higher and lower scientific-thinking ability. Item difficulty values, constrained to the 0.30-0.70 range during selection, ensured a test of moderate and appropriately varied difficulty rather than one weighted towards very easy or very difficult items.

The overall internal-consistency reliability ($KR-20 = 0.88$) and test-retest reliability ($r = 0.84$) both exceed accepted thresholds for cognitive-ability instruments, indicating that the test produces stable measurements both across items and across time. Component-wise $KR-20$ coefficients (ranging from 0.75 for Proportional and Correlational Reasoning to 0.82 for Controlling Variables) were somewhat lower than the overall coefficient, consistent with the smaller number of items (10) contributing to each component subscale, but remain within an acceptable range for research and diagnostic use. The comparatively higher reliability of the Controlling Variables component may reflect its more procedural, rule-based nature relative to the more abstractly demanding Proportional and Correlational Reasoning component, a pattern broadly consistent with Piagetian accounts of the differential difficulty of various formal-operational sub-skills.

The positive and significant criterion validity coefficient ($r = 0.61$, $p < .01$) obtained through correlation with science-subject achievement scores provides meaningful support for the test's validity, while its moderate rather than very high magnitude confirms that the Scientific Thinking Test captures a reasoning-based construct that is related to, but distinct from, content-based science achievement, consistent with the theoretical distinction between scientific reasoning and scientific knowledge that motivated the present study. The moderate inter-component correlations (0.33 to 0.52), combined with each component's stronger correlation with the total score than with other individual components, provide reasonable support for the five-component structure proposed,



while confirming that the components represent meaningfully related facets of a broader, unified scientific-thinking construct.

7. Major Findings

- A 50-item Scientific Thinking Test, comprising five components (Observation and Classification; Hypothesis Formulation and Deductive Reasoning; Controlling Variables; Proportional and Correlational Reasoning; and Data Interpretation and Inference), was successfully constructed and validated.
- Expert content validation yielded an average Content Validity Index of 0.90, indicating strong content validity.
- Item analysis on the try-out sample (N = 150) retained items with a mean discrimination index of 0.43 and difficulty values between 0.30 and 0.70, indicating good item quality.
- The final test demonstrated high internal-consistency reliability (KR-20 = 0.88) and test-retest reliability ($r = 0.84$).
- Criterion validity, established through correlation with science-subject achievement scores, was positive, moderate, and statistically significant ($r = 0.61$, $p < .01$).
- Percentile norms, grade-wise norms, and a five-category descriptive classification were successfully developed on a standardization sample of 650 secondary school students.

8. Educational Implications

- The standardized Scientific Thinking Test can be used by science teachers and curriculum evaluators to assess whether inquiry-based and activity-based pedagogy produces measurable gains in students' scientific reasoning ability, independent of content recall.



- Component-wise scores can help teachers identify specific reasoning sub-skills (for example, controlling variables or correlational reasoning) in which individual students or entire classes may require targeted instructional support.
- Curriculum planners and textbook developers can use aggregate findings from the test to identify which scientific-reasoning sub-skills are being adequately developed by current science curricula and which require greater instructional emphasis.
- The test provides researchers with a validated instrument for examining the relationship between scientific thinking and variables such as gender, educational stream, school area, intelligence, and academic achievement in future studies.

9. Conclusion

The present study successfully constructed and standardized a Scientific Thinking Test for secondary school students, addressing a notable gap in Indian educational-psychological instrumentation. Through a systematic process of item generation grounded in established theories of scientific reasoning, expert content validation, try-out item analysis, reliability estimation, criterion and construct validity assessment, and norm development, the resulting 50-item test was found to possess satisfactory psychometric properties. The test offers a contextually appropriate, curriculum-relevant, and psychometrically sound tool for assessing the scientific-reasoning ability of Indian secondary school students, distinct from conventional content-based science achievement testing, with clear applications in educational research, curriculum evaluation, and classroom-level diagnostic use.

References

- Dewey, J. (1910). *How we think*. D.C. Heath.
- Ebel, R. L. (1979). *Essentials of educational measurement* (3rd ed.). Prentice-Hall.



- Government of India. (1950). The Constitution of India, Article 51-A(h): Fundamental Duties.
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12(1), 1–48.
- Kuder, G. F., & Richardson, M. W. (1937). The theory of estimation of test reliability. *Psychometrika*, 2(3), 151–160.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575.
- Lawson, A. E. (1978). The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching*, 15(1), 11–24.
- National Council of Educational Research and Training. (2005). National curriculum framework 2005. NCERT.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.
- Piaget, J. (1970). *Genetic epistemology*. Columbia University Press.
- Tobin, K. G., & Capie, W. (1981). The development and validation of a group test of logical thinking. *Educational and Psychological Measurement*, 41(2), 413–423.
- Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, 27(2), 172–223.