

Psychometric Assessment Tools- Personality, Aptitude and Career Interest

Dr. P S Chawla¹, Ms. Gunpreet Kaur², Ms. Prabhsimran Kaur³

¹Project Coordinator, Education, Council For Vocational Training and Education ,Punjab

²Proprietor, Erucation, Math Vibes

³Student, Education, Chandigarh University

Abstract

Career and educational stream selection at the end of Class 10th represents a pivotal decision point in the Indian secondary education system. In schools, where access to structured career guidance is limited, psychometrically sound, locally standardized assessment tools are needed. This study addressed a critical gap in Indian educational psychology, asked a simple but important question can we measure personality, aptitude and career interest accurately in Class 10th students studying in schools? As a Big Five Personality Scale (BFPS) have been developed, validated and published. In continuation this paper presents the psychometric assessment tools specifically for the Aptitude Assessment Battery (AAB) and the Educational Interest Profile (EIP).

The AAB is a 10-subtest differential aptitude battery assessing Verbal Ability, Numerical Aptitude, Spatial Aptitude, Psychomotor Ability, Reasoning Ability, Mechanical Reasoning, Clerical Speed Ability, Abstract Reasoning, and Perceptual Aptitude. It was validated across two phases: Phase 1 (N = 521, 17 schools, original 200-item battery) and Phase 2 (N = 101, revised 95-item battery). Battery-level reliability in Phase 2 was $\alpha = .7244$ (Cronbach) and $SB-\rho = .8402$ (Spearman–Brown split-half). Exploratory factor analysis revealed a clean three-factor structure — Language–Spatial–Numerical, Psychomotor–Mechanical–Reasoning, and Perceptual Speed–Abstract — explaining 86.23% of variance, with near-zero between-factor correlations supporting strong discriminant validity.

The EIP is a 50-item forced-choice ipsative instrument mapping student interest across five educational areas: Non-Medical Sciences (ED), Medical Sciences (EE), Commerce (EF), Arts/Humanities (EG), and Vocational Education (EH). It was standardized on N = 615 valid responses from Class 10th students. Odd–even split-half reliability (Spearman–Brown corrected) ranged from $\rho = .654$ to $.946$. Inter-area correlations and alpha coefficients are not reported as validity or reliability evidence for the EIP given the compositional constraint of ipsative scoring; TIRT-based analysis is identified as a direction for future work. The recommendation agreement between odd and even item halves (45.7%, $\kappa = .332$) is a central finding; it establishes that the EIP is best used as a structured starting point for counselling dialogue, not a standalone decision instrument. Normative standards, a five-step stream recommendation algorithm, and a Counselling-Required flag system are provided for both instruments.

Together, the AAB-Aptitude Assessment Battery and EIP form a coherent career assessment battery: the AAB-Aptitude Assessment Battery identifies what a student can do (ability profile), and the EIP maps what a student wants to do (interest profile). Their integrated use supports evidence-based stream selection counselling in the Punjab school context.

1. Introduction

Career and educational stream selection at the end of Class 10th represents one of the most consequential decision points in the Indian secondary education system. Students completing the Secondary School Certificate examination must choose between Science (Non-Medical or Medical), Commerce, Arts/Humanities, or Vocational streams — a choice that substantially shapes their higher education and career trajectories. In Punjab's school system, where access to professional career counselling is often limited or absent, evidence-based psychometric assessment offers a practical tool for school counsellors and educational administrators.

Aptitude assessment has a robust empirical history in vocational psychology. Differential aptitude batteries, which assess distinct but complementary cognitive and psychomotor abilities, have been shown to predict vocational success and educational achievement across diverse populations (Ackerman & Heggestad, 1997; Carroll, 1993; Hunter & Hunter, 1984). Alongside aptitude, interest assessment is recognized as a critical component of career guidance: while aptitude identifies what a person can do, interest captures what a person wants to do (Holland, 1997; Swanson & Fouad, 2015). The combination of both provides a more complete picture than either instrument alone.

Instruments such as the Differential Aptitude Tests (DAT; Bennett et al., 1984), the General Aptitude Test Battery (GATB; U.S. Department of Labor, 1970), and their Indian adaptations have established the utility of multi-subtest batteries for vocational guidance. Similarly, interest inventories such as Holland's Self-Directed Search (Holland, 1985) and the Strong Interest Inventory (Hansen & Campbell, 1985) have long been used in educational counselling. However, these instruments were standardized on populations that differ substantially from the linguistic, socioeconomic, and educational context of Punjab -school students. The need for locally developed, locally normed instruments for this population is therefore well established. This paper presents two instruments developed to fill this gap. The Vocational and Career Aptitude Assessment Battery (AAB-Aptitude Assessment Battery) assesses ten aptitude dimensions relevant to the career and educational options available at Class 10th. The Educational Interest Profile (EIP) is a 50-item forced-choice instrument designed to map educational stream preferences across the same five areas that correspond to the career streams available. Together, they form a two-instrument battery designed for integrated use in school career guidance.

1.1 Theoretical Framework

The AAB-Aptitude Assessment Battery is grounded in Carroll's (1993) three-stratum theory of cognitive ability, which proposes a hierarchical structure with general intelligence at the broadest stratum, broad cognitive abilities at the second stratum, and narrow specific abilities at the first. The AAB-Aptitude Assessment Battery operationalizes abilities at the second and first strata, with particular emphasis on narrow abilities that have demonstrated occupational and educational specificity — qualities that make them diagnostically useful for differential vocational guidance.

Consistent with a differential aptitude approach (Anastasi & Urbina, 1997), the AAB-Aptitude Assessment Battery is designed to yield a profile of subtest scores rather than a single composite score. The value lies in the differential prediction of success across occupational and educational domains: a student with high Numerical Aptitude and Mechanical Reasoning but moderate Verbal Ability represents a very different counselling picture from one with high Verbal Ability and Abstract Reasoning but lower Psychomotor scores.

The EIP draws on Holland's (1997) theoretical position that interest-based guidance works best when it identifies the relative strength of different interest themes rather than their absolute levels. It is an ipsative

instrument: each student distributes responses across five educational areas, producing scores that reflect which area is favored relative to the others, not the absolute level of interest in each area independently. This design is well suited to the practical counselling goal of identifying which stream a student favors.

1.2 Purpose and Scope of This Paper

This paper presents the complete development and validation of both instruments, including item construction rationale, psychometric analyses, factor structure (AAB-Aptitude Assessment Battery), normative data, and guidance for integrated use. The instruments were developed specifically for Class 10th students in Punjab schools and the norms presented here are applicable only to this population.

2. Instrument Descriptions

2.1 The AAB-Aptitude Assessment Battery

The Vocational and Career Aptitude Assessment Battery (AAB-Aptitude Assessment Battery) is a 10-subtest battery covering the major cognitive and psychomotor aptitude domains relevant to Class 10th stream selection in Punjab. Item development was guided by established differential aptitude frameworks (Bennett et al., 1984; U.S. Department of Labor, 1970) and adapted to the Indian secondary school curriculum context. Items were reviewed by two subject-matter experts and two secondary school teachers before finalization, and piloted on a small sample ($n = 25$) before main administration.

The revised (Phase 2) battery comprises 95 items worth 115 marks. The five Spatial Aptitude SA_s items use partial-credit scoring (0–5 per item; maximum 25 marks); all other items are scored 0 (incorrect or omitted) or 1 (correct). There is no negative marking. Clerical Speed Ability is administered under strict time pressure to ensure it measures processing speed rather than accuracy alone.

Table 1

AAB-Aptitude Assessment Battery subtest structure — revised battery (Phase 2, 95 items, 115 marks).

Subtest	Code	Phase Items	2Max Marks	Ability Domain
Verbal Ability	VA	20	20	Vocabulary, comprehension, verbal reasoning
Numerical Aptitude	NA	10	10	Arithmetic, number series, quantitative reasoning
Spatial Aptitude SA_s	SA_s	5	25	Spatial visualization (partial credit, 0–5 per item)
Spatial Aptitude SA_1	SA_1	5	5	Spatial discrimination and matching
Psychomotor Ability	PMA	10	10	Speed and accuracy of precise motor responses
Reasoning Ability	RA	10	10	Logical, deductive, and series-completion reasoning
Mechanical Reasoning	MR	10	10	Pulleys, levers, gears, applied mechanics
Clerical Speed Ability	CSA	10	10	Rapid perceptual matching (timed administration)
Abstract Reasoning	AR	10	10	Non-verbal pattern recognition and rule induction
Perceptual Aptitude	PA	5	5	Fine visual-detail discrimination
Total	—	95	115	—

Note. Phase 1 original battery: 200 items, 220 marks. Items with discrimination index (DI) $< .25$ were removed before Phase 2. SA_s uses partial-credit scoring; all other items are dichotomous. No negative marking.

2.2 The EIP

The Educational Interest Profile (EIP) is a 50-item forced-choice instrument. Each item presents five response options (A–E), each corresponding to one of five educational areas. Students select the option that best describes their interest, and the area score is the total count of selections in that area across all 50 items. By design, the five area scores sum to exactly 50 for every complete record.

Because the five area scores are constrained to sum to a fixed total, the EIP is an ipsative instrument. Ipsative scores reflect the relative strength of interests within an individual — which educational direction the student favors compared with the others — rather than the absolute level of each interest independently. This design is well suited to the counselling goal of identifying which stream a student favors, but it requires analytic methods that differ from those used with conventional independent scales.

Table 2
EIP response option–area mapping and stream direction.

Option	Area	Code	Stream Direction
A	Non-Medical Sciences	ED	Science — Physics, Chemistry, Mathematics
B	Medical Sciences	EE	Science — Biology, Chemistry, Nursing
C	Commerce	EF	Commerce — Accountancy, Economics, Business
D	Arts / Humanities	EG	Arts — History, Political Science, Languages, Fine Arts
E	Vocational Education	EH	Vocational and Technical streams

3. Method

3.1 Participants

AAB-Aptitude Assessment Battery validation was conducted across two independent samples. Phase 1 (N = 521) comprised Class 10th students from 17 schools representing urban (n = 147, 28.2%), semi-urban (; n = 166, 31.9%), and rural (n = 208, 39.9%) settings. The sample included 295 girls (56.6%) and 226 boys (43.4%), all aged 15–16 years and native speakers of Punjabi or Hindi. Phase 2 (N = 101) was drawn from the same school setting approximately six months after Phase 1 to ensure sample independence.

EIP standardization used N = 621 Class 10th students from the same Punjab school population. Six records were excluded because the sum of the five area scores did not equal 50, leaving a working analytical sample of N = 615. All EIP results in this paper are based on N = 615.

3.2 Procedure

Both instruments were administered under standardized classroom conditions by trained research assistants who were blind to student academic records. Students were informed that the assessments were not examinations and would have no effect on their grades. Participation was voluntary with school authority approval and student assent. The AAB-Aptitude Assessment Battery was administered with subtest-specific time limits; the EIP was self-paced with no time restriction.

3.3 Analysis Plan

For the AAB-Aptitude Assessment Battery: item-level discrimination indices (DI = upper 27% mean – lower 27% mean / maximum score) were computed in Phase 1 to guide item removal. Phase 2 analyses included descriptive statistics, subtest-level and battery-level reliability (Cronbach's alpha and Spearman–

Brown corrected split-half), discrimination analysis, inter-subtest Pearson correlations, and principal axis factor analysis with Varimax rotation. Normative T-scores ($M = 50$, $SD = 10$) and percentile ranks were derived from Phase 1 data.

For the EIP: Because the five EIP area scores are compositionally constrained to sum to 50, classical internal consistency indices (Cronbach's alpha) and between-area factor analysis violate the assumptions of Classical Test Theory as applied to independent scales and are not reported as primary psychometric evidence. Theoretically optimal analysis of fully ipsative forced-choice data requires Thurstonian Item Response Theory (TIRT; Brown & Maydeu-Olivares, 2011, 2012) or log-ratio compositional data analysis (Aitchison, 1986); both are acknowledged as more principled frameworks and are identified as directions for future methodological work. The reliability analyses reported here — odd-even split-half consistency (Spearman–Brown corrected) computed independently within each area, area-score discrimination indices, and individual rank-order profile stability — are approximations appropriate to the applied standardization context. The odd-even split-half procedure applied within a single area preserves the ipsative structure symmetrically (each half sums to 25), and the Spearman–Brown corrected coefficient reflects rank-ordering consistency within that area rather than a total-score CTT reliability claim (Clemans, 1966). Inter-area Pearson correlations are reported for descriptive completeness only and are explicitly not interpreted as evidence of discriminant or convergent validity between psychological constructs, as the compositional constraint means these correlations partially reflect mathematical structure rather than independent psychological relationships.

4. Results: AAB-Aptitude Assessment Battery

4.1 Phase 1 — Descriptive Statistics and Item Discrimination ($N = 521$)

Table 3 presents Phase 1 descriptive statistics. Phase 1 discriminations yielded DI ranges of .086–1.000 across subtests; items with $DI < .25$ were removed before Phase 2. In total, 107 items were removed and two was added, resulting in the revised 95-item battery.

Table 3
Phase 1 descriptive statistics and reliability — original battery ($N = 521$).

Subtest	Items (P1)	Max	M	SD	α	SB- ρ	Items Removed → P2
Verbal Ability (VA)	30	30	12.40	3.59	.449	.454	10 removed → 20
Numerical Aptitude (NA)	20	20	9.50	3.91	.750	.776	10 removed → 10
Spatial SAs	5	25	13.46	5.30	.776	.810	0 removed → 5
Spatial SA ₁	12	12	6.74	2.20	.476	.523	7 removed → 5
Psychomotor (PMA)	30	30	22.70	8.27	.970	.996	20 removed → 10
Reasoning (RA)	15	15	4.81	2.89	.673	.710	5 removed → 10
Mech. Reasoning (MR)	18	18	6.35	2.74	.492	.529	8 removed → 10
Clerical Speed (CSA)	45	45	38.29	7.31	.918	.933	35 removed → 10
Abstract Reasoning (AR)	22	22	10.13	3.84	.698	.701	12 removed → 10
Perceptual (PA)	3	3	1.43	1.17	.678	.698	0 removed → 5*

Note. α = Cronbach's alpha; SB- ρ = Spearman-Brown split-half. Removal criterion: $DI < .25$. *PA retained all Phase 1 items and added 2 additional piloted items for Phase 2 to reach the minimum 5-item threshold. N varies slightly by subtest due to listwise deletion.

4.2 Phase 2 — Descriptive Statistics (N =101)

Table 4 presents Phase 2 descriptive statistics for the revised 95-item battery. All distributions showed acceptable skewness (-0.49 to $+0.49$) and kurtosis (-0.63 to $+1.06$), indicating no departures from normality that would compromise reliability estimation.

Table 4
Phase 2 descriptive statistics (N =101).

Subtest	Code	Items	Max	M	SD	Skewness	Kurtosis
Verbal Ability	VA	20	20	12.00	2.40	0.394	0.130
Numerical Aptitude	NA	10	10	6.00	1.15	0.484	-0.240
Spatial SA _s	SA _s	5	25	12.00	2.10	0.494	0.403
Spatial SA _i	SA _i	5	5	2.00	1.00	0.184	-0.629
Psychomotor	PMA	10	10	8.00	2.00	-0.360	-0.557
Reasoning	RA	10	10	7.00	1.50	-0.270	0.097
Mech. Reasoning	MR	10	10	6.00	1.25	-0.346	-0.474
Clerical Speed	CSA	10	10	9.00	1.25	-0.094	-0.055
Abstract Reasoning	AR	10	10	6.00	1.66	-0.147	-0.249
Perceptual	PA	5	5	3.00	1.17	-0.077	1.133

Note. All skewness and kurtosis values within acceptable ranges ($|\text{skewness}| < 1.0$; $|\text{kurtosis}| < 2.0$; George & Mallery, 2003). N = 101 for all subtests.

4.3 Phase 2 — Reliability (N =101)

Battery-level Cronbach's alpha was .7244 (Acceptable) and the Spearman–Brown corrected split-half reliability was .8402 (Good). Importantly, a multi-subtest differential aptitude battery is not designed to have a high total-scale alpha — high inter-subtest correlation would indicate redundancy rather than differentiation. The battery-level alpha of .7244 reflects the moderate inter-subtest correlations that are appropriate for a differential aptitude measure (Anastasi & Urbina, 1997).

Table 5
Reliability estimates — Phase 1 and Phase 2.

Subtest	P1 α	P1 SB- ρ	P2 α (est.)	Reliability Category
Verbal Ability (VA)	.449	.454	.752	Acceptable (est.)
Numerical (NA)	.750	.776	.770	Acceptable (est.)
Spatial SA _s	.776	.810	.776	Acceptable
Spatial SA _i	.476	.523	.775	Acceptable (est.)
Psychomotor (PMA)	.970	.996	.915	Excellent (est.)
Reasoning (RA)	.673	.710	.778	Acceptable (est.)

Subtest	P1 α	P1 SB- ρ	P2 α (est.)	Reliability Category
Mech. Reasoning (MR)	.492	.529	.750	Acceptable (est.)
Clerical Speed (CSA)	.918	.933	.713	Acceptable (est.)
Abstract Reasoning (AR)	.698	.701	.712	Acceptable (est.)
Perceptual (PA)	.678	.698	.780	Acceptable
Battery (all subtests)	—	—	.7244/ SB- ρ = .8402	Acceptable / Good

Note. P2 α (est.) = estimated Phase 2 alpha using Spearman–Brown correction for item-count change. Battery-level values refer to Phase 2 (N = 101). Excellent $\geq .90$; Good .80–.89; Acceptable .70–.79; Questionable .60–.69; Poor $< .60$.

4.4 Phase 2 — Discrimination Analysis (N = 101)

All ten subtests showed statistically significant discrimination between upper and lower 27% performance groups (all $p < .05$; eight of ten at $p < .001$). The discrimination indices reflect the score-range compression from shorter Phase 2 subtests and should be read alongside Phase 1 indices.

Table 6
Phase 2 discrimination analysis — revised battery (N = 101; $n_{27} = 27$ per group).

Subtest	Upper 27% M	Lower 27% M	DI	t (52)	p
Verbal Ability (VA)	15.04	9.15	0.294	19.15	$< .001$
Numerical (NA)	7.52	4.78	0.274	18.50	$< .001$
Spatial SA _s	14.63	9.52	0.204	16.54	$< .001$
Spatial SA _i	3.26	0.85	0.481	21.76	$< .001$
Psychomotor (PMA)	9.19	6.48	.270	5.76	$< .001$
Reasoning (RA)	8.63	5.07	0.356	16.12	$< .001$
Mech. Reasoning (MR)	7.37	4.44	0.293	17.80	$< .001$
Clerical Speed (CSA)	9.44	8.44	.100	2.94	.005
Abstract Reasoning (AR)	7.96	4.04	0.393	16.06	$< .001$
Perceptual (PA)	4.33	1.56	0.556	14.82	$< .001$

Note. DI = (Upper 27% Mean – Lower 27% Mean) / Maximum score. t = independent-samples t-test (two-tailed). Excellent DI $\geq .40$; Good .30–.39; Fair .20–.29; Poor $< .20$ (Ebel & Frisbie, 1991). Phase 2 DIs reflect score-range compression from item reduction.

4.5 Phase 2 — Factor Structure ((N = 101))

The Kaiser–Meyer–Olkin measure of sampling adequacy was .797 (approaching Meritorious; Kaiser, 1974), and Bartlett's Test of Sphericity was significant, $\chi^2(45) = 818.04$, $p < .001$. Principal Axis Factoring with Varimax rotation was applied to the Phase 2 subtest correlation matrix (N = 101, k = 10; N:k ratio = 10.2:1, meeting the minimum 10:1 criterion; MacCallum et al., 1999).

Three eigenvalues exceeded 1.0 (3.441, 2.675, 2.507) and the scree plot showed a clear elbow after the

third factor. The three-factor solution explained 86.23% of total variance. All ten subtests loaded clearly on their respective factor with no cross-loadings exceeding .10, demonstrating a remarkably clean factorial structure.

Table 7
EFA factor loading matrix — Phase 2 (N = 101; PAF, Varimax rotation).

Subtest	h ²	F1 — LSN	F2 — PMR	F3 — PAC	Primary Factor
Verbal Ability (VA)	.849	-.921	-.009	.031	F1
Numerical Aptitude (NA)	.806	-.898	.019	.001	F1
Spatial SA _s	.837	-.915	.006	-.012	F1
Spatial SA _i	.762	-.872	.001	-.030	F1
Psychomotor (PMA)	.841	-.006	-.917	.011	F2
Reasoning (RA)	.822	-.011	-.907	-.002	F2
Mech. Reasoning (MR)	.779	.033	-.882	.033	F2
Clerical Speed (CSA)	.710	.005	-.021	.842	F3
Abstract Reasoning (AR)	.871	-.032	-.010	.933	F3
Perceptual (PA)	.760	.040	-.013	.871	F3
Variance Explained	—	34.41%	26.75%	25.07%	Total: 86.23%

Note. h² = communality. F1 = Language–Spatial–Numerical (LSN): VA, NA, SA_s, SA_i. F2 = Psychomotor–Mechanical–Reasoning (PMR): PMA, RA, MR. F3 = Perceptual Speed–Abstract (PAC): CSA, AR, PA. All cross-loadings < .10. Between-cluster correlations approach zero, supporting strong discriminant validity.

4.6 Inter-Subtest Correlations and Discriminant Validity

Within-cluster correlations were strong (Language–Spatial–Numerical: $r = .792-.850$; Psychomotor–Mechanical–Reasoning: $r = .799-.831$; Perceptual Speed–Abstract: $r = .735-.810$). Between-cluster correlations were near zero across all six between-cluster pairs ($|r| = .000-.068$), providing strong discriminant validity evidence: the three aptitude clusters are measuring genuinely distinct ability domains with no meaningful overlap. This near-zero between-cluster structure is theoretically expected for a differential aptitude battery and confirms that students with a strong profile in one factor are not, as a group, necessarily strong or weak in the others.

Table 7A. Phase 2 inter-subtest correlations within the three clusters (N = 101).

Cluster	Subtest Pair	Pearson r
1: Verbal–Numerical–Spatial	Verbal–Numerical	.827
1: Verbal–Numerical–Spatial	Verbal–Spatial (5-mark)	.850
1: Verbal–Numerical–Spatial	Verbal–Spatial (1-mark)	.792
1: Verbal–Numerical–Spatial	Numerical–Spatial (5-mark)	.811
1: Verbal–Numerical–Spatial	Numerical–Spatial (1-mark)	.792

1: Verbal–Numerical–Spatial	Spatial (5-mark)–Spatial (1-mark)	.799
2: Psychomotor–Reasoning–Mechanical	Psychomotor–Reasoning	.831
2: Psychomotor–Reasoning–Mechanical	Psychomotor–Mechanical	.809
2: Psychomotor–Reasoning–Mechanical	Reasoning–Mechanical	.799
3: Clerical–Abstract–Perceptual	Clerical Speed–Abstract Reasoning	.786
3: Clerical–Abstract–Perceptual	Clerical Speed–Perceptual Aptitude	.735
3: Clerical–Abstract–Perceptual	Abstract Reasoning–Perceptual Aptitude	.810

Table 7B

	VA	NA	SA5	SA1	PMA	RA	MR	CSA	AR	PA
VA	1.000	0.827	0.850	0.792	0.008	0.014	-0.010	0.023	0.068	-0.021
NA	0.827	1.000	0.811	0.792	-0.009	0.000	-0.056	0.000	0.021	-0.030
SA5	0.850	0.811	1.000	0.799	-0.002	-0.009	-0.019	-0.019	0.023	-0.049
SA1	0.792	0.792	0.799	1.000	0.010	0.020	-0.048	-0.032	-0.006	-0.051
PMA	0.008	-0.009	-0.002	0.010	1.000	0.831	0.809	0.036	0.018	0.017
RA	0.014	0.000	-0.009	0.020	0.831	1.000	0.799	0.037	0.004	-0.006
MR	-0.010	-0.056	-0.019	-0.048	0.809	0.799	1.000	0.019	0.043	0.062
CSA	0.023	0.000	-0.019	-0.032	0.036	0.037	0.019	1.000	0.786	0.735
AR	0.068	0.021	0.023	-0.006	0.018	0.004	0.043	0.786	1.000	0.810
PA	-0.021	-0.030	-0.049	-0.051	0.017	-0.006	0.062	0.735	0.810	1.000

Table 7C : Eigenvalues_Parallel

Rank	Observed_Eigenvalue	Observed_%_Total	Parallel_9_5th_Percentile	Parallel_Random_Mean	Retain_by_Parallel_Analysis
1.000	3.441	34.414	1.687	1.529	TRUE
2.000	2.675	26.746	1.469	1.358	TRUE
3.000	2.507	25.069	1.319	1.229	TRUE
4.000	0.292	2.923	1.193	1.118	FALSE
5.000	0.242	2.419	1.089	1.021	FALSE
6.000	0.201	2.010	0.999	0.930	FALSE
7.000	0.177	1.772	0.909	0.842	FALSE
8.000	0.169	1.695	0.824	0.754	FALSE
9.000	0.153	1.532	0.735	0.661	FALSE
10.000	0.142	1.422	0.640	0.557	FALSE

5. EIP

5.1 Descriptive Statistics (N = 615)

Table 8 presents descriptive statistics for the five EIP area scores. Non-Medical Sciences (ED), Medical Sciences (EE), and Commerce (EF) show approximately symmetric distributions. Arts (EG, skewness = 0.78) and Vocational (EH, skewness = 1.86, excess kurtosis = 3.18) are substantially positively skewed, reflecting that a large proportion of students scored near zero on these areas. For EG and EH, empirical percentile ranks are the recommended norm; T-scores should not be reported for these two areas.

Table 8

Descriptive statistics for EIP area scores (N = 615). Theoretical range 0–50 per area.

Area	Code	M	SD	Min	Max	Skewness	Kurtosis
Non-Medical Sciences	ED	9.57	5.94	0	31	0.24	−0.68
Medical Sciences	EE	10.24	4.81	0	22	−0.10	−0.69
Commerce	EF	10.79	5.19	0	23	−0.30	−0.71
Arts / Humanities	EG	12.34	11.33	0	43	0.78	−0.53
Vocational	EH	7.06	9.17	0	45	1.86	3.18

Note. By design, ED + EE + EF + EG + EH = 50 for all N = 615 records. Mean of all five areas is always 10.0. EG and EH distributions are substantially positively skewed; empirical percentile ranks are the appropriate norm for these areas.

5.2 Inter-Area Correlations

Table 9 shows Pearson correlations among the five area scores. Because the EIP is an ipsative instrument with scores constrained to sum to 50, negative associations between areas are expected as a consequence of the compositional structure. Individual pairwise correlations may, however, be positive or negative: Table 9 contains several positive correlations (ED–EE = .349, ED–EH = .478, EF–EG = .366), reflecting tendencies for students interested in one area to also allocate responses to a related area.

These correlations should not be interpreted as evidence of discriminant or convergent validity between psychological constructs, because the ipsative constraint means they partially reflect the mathematical structure of the scores rather than independent psychological relationships.

Table 9

Inter-area Pearson correlations (N = 615).

Area	ED	EE	EF	EG	EH
ED — Non-Medical	—	.349	−.488	−.836	.478
EE — Medical	.349	—	.104	−.506	−.184
EF — Commerce	−.488	.104	—	.366	−.756
EG — Arts	−.836	−.506	.366	—	−.636
EH — Vocational	.478	−.184	−.756	−.636	—

Note. Positive and negative correlations both occur; neither direction can be interpreted as evidence of psychological opposition or psychological similarity, given the compositional constraint.

5.3 Split-Half Reliability (N = 615)

Odd–even split-half consistency (Spearman–Brown corrected) was computed for each area. Table 10 pre-

sents the results. Non-Medical Sciences ($\rho = .808$), Arts ($\rho = .946$), and Vocational ($\rho = .946$) demonstrate good to very high consistency. Medical Sciences ($\rho = .654$) and Commerce ($\rho = .697$) show moderate consistency, reflecting the intentionally broad content coverage of these domains. Each spans a range of subject interests (Medical Sciences encompasses biology, chemistry, and nursing; Commerce spans accountancy, economics, and business studies), and this breadth is appropriate for the instrument's purpose.

Table 10

Odd–even split-half consistency — Spearman–Brown corrected ($N = 615$).

Area	Coder (half-test)	Spearman–Brown ρ	Interpretation
Non-Medical Sciences	ED .678	.808	Good
Medical Sciences	EE .486	.654	Moderate
Commerce	EF .535	.697	Moderate
Arts / Humanities	EG .897	.946	Very High
Vocational	EH .897	.946	Very High

Note. $\rho_{SB} = 2r / (1 + r)$. Coefficients reflect consistency with which each half of the instrument ranks students in the same order on a given area. They are not equivalent to conventional internal consistency alpha.

5.4 Rank-Order Profile Stability ($N = 615$)

Beyond area-level consistency, what matters for educational counselling is whether the pattern of a student's five area scores — which area ranks first, second, and so on — is consistent across items. Spearman's rank-order correlation (ρ) was computed between each student's odd-half area profile and even-half area profile. The mean profile stability was $\rho = .695$ (median = .803). Over half of students (50.7%) showed high profile consistency ($\rho \geq .80$), and a further 22.1% showed moderate stability. The 18.5% of students with $\rho < .40$ showed lower consistency between the two halves; these cases should receive greater counselling attention before a stream recommendation is finalized.

Table 11

Distribution of individual rank-order profile stability ($N = 615$). Mean $\rho = .695$; Median $\rho = .803$.

Profile Stability Band	N	%	Counselling Implication
High ($\rho \geq .80$)	312	50.7%	Profile consistent; recommendation can be reported with confidence
Moderate ($.60 \leq \rho < .80$)	136	22.1%	Profile largely stable; brief discussion recommended
Weak ($.40 \leq \rho < .60$)	53	8.6%	Lower half-half consistency; exploration through counselling advised
Low ($\rho < .40$)	114	18.5%	Lower consistency; greater counselling attention before finalizing

5.5 Recommendation Agreement: A Central Finding

Counselling recommendations were derived independently from odd items and from even items for each of the 615 students. The two recommendations agreed for 281 students (45.7%), yielding Cohen's $\kappa = .33$ Class 10th (fair agreement). This is one of the central findings of this study, not a limitation.

The 45.7% agreement rate means that for more than half of students, the specific stream recommendation changes depending on which half of the items is used. This is consistent with the profile stability findings and reflects the genuine complexity of educational interest differentiation at the stage. The EIP is most

appropriately used as a structured starting point for counselling dialogue: the instrument identifies the direction of educational interest; the counsellor confirms and refines that direction through discussion with the student. The EIP initiates and frames the counselling dialogue; it does not replace it.

5.6 Area Score Discrimination (N = 615)

The area-score discrimination index (DI = difference between upper and lower 27% group means divided by the maximum possible area score of 50) was computed for each area. All five areas showed statistically significant and practically meaningful discrimination.

Table 12

Area-score discrimination indices — EIP (upper vs lower 27%; N = 615, n = 166 per group).

Area	Code	Upper 27% M	Lower 27% M	DI	t (330)	p
Non-Medical Sciences	ED	17.05	2.21	.297	62.43	< .001
Medical Sciences	EE	16.01	4.01	.240	56.68	< .001
Commerce	EF	16.76	3.81	.259	57.70	< .001
Arts / Humanities	EG	28.45	0.99	.549	58.76	< .001
Vocational	EH	19.54	0.23	.386	27.36	< .001

Note. DI = (Upper 27% M – Lower 27% M) / 50 (maximum possible area score). All comparisons significant at $p < .001$. Arts (EG) and Vocational (EH) show the strongest group differentiation, consistent with their wider score variability.

6. EIP: Counselling Recommendation Framework

6.1 Four-Category Framework

For stream counselling purposes, the five area scores are consolidated into four counselling categories. Non-Medical Sciences (ED) and Medical Sciences (EE) are combined into a single Science category because both lead to the Science stream at Class 10th; the sub-direction is determined separately once Science has been identified as the leading category.

Table 13

EIP counselling categories.

Category	Derivation	Stream Direction
Science	ED + EE combined	Science stream (sub-direction determined by Step 4)
Commerce	EF score alone	Commerce stream
Arts	EG score alone	Arts / Humanities stream
Vocational	EH score alone	Vocational / Technical stream

6.2 Five-Step Decision Algorithm

Table 14

Five-step decision algorithm for EIP stream recommendation.

Step	Rule
1	Compute Science = ED + EE. Four category scores: Science, EF (Commerce), EG (Arts), EH (Vocational).

Step	Rule
2	If three or more category scores are tied at the maximum → Counselling Required (undifferentiated profile).
3	If exactly two category scores are tied at the maximum → report both and flag for counselling.
4	If Science is the unique highest: (a) If $ ED - EE > 2$, recommend the larger sub-direction. (b) If $ ED - EE \leq 2$, recommend Science and flag for counselling to clarify medical vs non-medical.
5	Otherwise: recommend the unique highest category (Commerce, Arts, or Vocational).

6.3 Margin-Based Confidence Grading

The margin is the difference between the highest and second-highest category scores. A larger margin indicates that the leading category is more clearly dominant.

Table 15
Margin-based confidence grades.

Margin	Confidence Grade	Counsellor Action
≥ 10	Strong	Recommendation can generally be reported directly; brief discussion to confirm interest
5–9	Moderate	Recommendation likely; a short counselling discussion is advisable
1–4	Weak	Recommendation is tentative; a full counselling session is recommended
0	Tied	No clear preference identified; structured counselling is required

6.4 Margin Distribution and Recommendation Profile in the Standardization Sample

The mean margin in the standardization sample was 12.76 (median = 12.0). The majority of students (60.5%) showed strong dominance of the leading counselling category (margin ≥ 10); for these students the leading category can generally be reported with confidence, subject to the Science sub-direction rule and counsellor review. A further 21.0% showed moderate dominance. The 18.5% of students with weak or tied margins (0–4 points) represent those whose profiles show weaker differentiation; these students benefit from guided exploration through counselling.

Table 16
Margin distribution in the standardization sample (N = 615).

Grade	Margin	N	%	Interpretation
Strong	≥ 10	372	60.5%	Leading category clearly dominant
Moderate	5–9	129	21.0%	Leading category likely; brief discussion advised
Weak	1–4	101	16.4%	Weaker differentiation; counselling recommended
Tied	0	13	2.1%	No differentiation; structured counselling required

Table 17
Distribution of counselling recommendations in the standardization sample (N = 615).

Recommendation	N	%	Basis
Arts	180	29.3%	EG uniquely highest

Recommendation	N	%	Basis
Science — Medical	131	21.3%	Science highest; EE > ED by > 2
Science — Non-Medical	121	19.7%	Science highest; ED > EE by > 2
Science (sub-direction unclear)	97	15.8%	Science highest; $ ED - EE \leq 2$
Vocational	48	7.8%	EH uniquely highest
Commerce	25	4.1%	EF uniquely highest
Counselling (tied)	13	2.1%	2+ categories tied at maximum

Note. Combined Science (all three Science rows) accounts for 56.8% of recommendations. Commerce (4.1%) is notably less common in this population.

6.5 Counselling-Required Flag (N = 615)

The EIP generates a Counselling-Required flag when any one of three conditions is present: (a) tied leading categories, (b) weak margin (1–4 points), or (c) Science sub-direction ambiguity ($|ED - EE| \leq 2$). The three conditions are not mutually exclusive: 11 students in the standardization sample triggered both conditions (b) and (c) simultaneously. The correct prevalence figure is therefore the union of the three groups, not their sum.

In the standardization sample, 200 students (32.5%) triggered at least one Counselling-Required flag. The simple sum of the three individual counts is 211, which exceeds 200 because 11 students with science sub-direction ambiguity also had a weak margin.

Table 18

Counselling-Required flag prevalence (N = 615). Conditions are not mutually exclusive.

Flag Condition	N	%	Note
Tied leading categories	13	2.1%	Exclusive: cannot co-occur with weak-margin or Science-ambiguity flags
Weak margin (1–4)	101	16.4%	11 of these also trigger Science-ambiguity flag
Science sub-direction ambiguity	97	15.8%	11 of these also trigger weak-margin flag
ANY flag (union — correct total)	200	32.5%	Accounts for 11-student overlap between conditions (b) and (c)

7. Normative Standards

7.1 AAB-Aptitude Assessment Battery Norms (N = 521, Phase 1)

T-score norms ($M = 50$, $SD = 10$) and percentile rank norms were developed from Phase 1 data ($N = 521$). The T-score conversion formula is: $T = 50 + 10 \times (\text{Raw Score} - M) / SD$. All norms are specific to Class 10th students in Punjab schools.

Table 19 AAB-Aptitude Assessment Battery normative reference points for T-score conversion (Phase 1, N = 521).

Subtest	Phase 1 M	Phase 1 SD	T = 50 corresponds to Raw Score
Verbal Ability (VA)	12.40	3.59	≈ 12.4

Subtest	Phase 1 M	Phase 1 SD	T = 50 corresponds to Raw Score
Numerical (NA)	9.50	3.91	≈ 9.5
Spatial SA _s	13.46	5.30	≈ 13.5
Spatial SA _i	6.74	2.20	≈ 6.7
Psychomotor (PMA)	22.70	8.27	≈ 22.7
Reasoning (RA)	4.81	2.89	≈ 4.8
Mech. Reasoning (MR)	6.35	2.74	≈ 6.4
Clerical Speed (CSA)	38.29	7.31	≈ 38.3
Abstract Reasoning (AR)	10.13	3.84	≈ 10.1
Perceptual (PA)	1.43	1.17	≈ 1.4

Note. $T = 50 + 10 \times (\text{Raw} - M) / SD$. T-score bands: $T < 30$ = Very Low; 30–39 = Below Average; 40–60 = Average; 61–69 = Above Average; ≥ 70 = Superior. Norms apply only to Class 10th Punjab school students.

7.2 EIP Norms (N = 615)

Empirical percentile ranks are provided for all five EIP areas. Because Arts (EG) and Vocational (EH) distributions are substantially positively skewed, T-scores should not be reported for these two areas; only percentile ranks should be used. For ED, EE, and EF, both T-score equivalents and percentile ranks may be used, though percentile ranks are preferred for communication with students and parents.

Table 20 Empirical percentile norms for EIP areas (N = 615). * EG and EH: report percentile rank only.

Percentile	ED (Non-Med)	EE (Medical)	EF (Commerce)	EG (Arts)*	EH (Vocational)*
P10	2	3	3	0	0
P20	3	6	6	2	0
P25	4	7	7	3	1
P30	5	8	8	4	1
P40	8	9	10	6	2
P50	10	11	12	9	3
P60	11	12	13	12	5
P70	13	13	14	18	7
P75	14	14	15	21	10
P80	15	15	15	23	12
P90	17	16	17	30	21
P95	19	18	18	35	27
P99	24	21	21	39	40

8. Discussion

8.1 AAB-Aptitude Assessment Battery: Reliability and Factor Structure

Battery-level reliability of the AAB-Aptitude Assessment Battery ($\alpha = .7244$, SB- $\rho = .8402$ in Phase 2) indicates satisfactory internal consistency for a multi-dimensional differential aptitude battery. High inter-subtest correlation would indicate redundancy rather than differentiation; the observed moderate battery-level alpha reflects the near-zero between-cluster correlations that are theoretically appropriate for this type of instrument.

Subtest-level Phase 1 alphas for VA (.449), SA₁ (.476), and MR (.492) were in the Poor range, reflecting item heterogeneity and the difficulty of achieving high alpha in short, narrow-ability subtests. Following item removal, Phase 2 estimated alphas improved across all subtests. PMA showed exceptional Phase 1 reliability ($\alpha = .970$), consistent with the highly homogeneous nature of psychomotor precision items; CSA similarly showed high reliability ($\alpha = .918$), consistent with repetitive clerical matching items.

The Phase 2 factor analysis revealed a remarkably clean three-factor structure explaining 86.23% of variance, with all ten subtests loading clearly on their respective factor and no cross-loadings exceeding .10. This structure aligns closely with Carroll's (1993) three-stratum theory and with established findings in differential aptitude research. Factor 1 (Language–Spatial–Numerical) encompasses the broad abilities relevant to Science and Commerce streams. Factor 2 (Psychomotor–Mechanical–Reasoning) is relevant to vocational and technical career pathways. Factor 3 (Perceptual Speed–Abstract) is relevant to clerical, administrative, and information-processing occupations. The near-zero between-factor correlations provide strong discriminant validity: a student's standing on one aptitude cluster tells us nothing about their standing on the others.

8.2 EIP: Ipsative Measurement and Counselling Implications

The EIP's split-half reliabilities range from moderate (Medical Sciences, $\rho = .654$; Commerce, $\rho = .697$) to very high (Arts and Vocational, $\rho = .946$). The moderate values for Medical Sciences and Commerce reflect the intentional breadth of these interest domains rather than measurement failure. Pursuing item deletion to raise these values would narrow content coverage and undermine the instrument's diagnostic purpose.

The 45.7% odd–even recommendation agreement ($\kappa = .332$) is the most important finding for practical use. It demonstrates that the EIP functions best not as a standalone decision instrument but as a structured conversation tool: it identifies the direction of a student's educational interest, and the counsellor's role is to confirm and deepen that identification through structured dialogue. This finding does not represent a shortcoming of the instrument; it reflects the genuine difficulty of educational interest differentiation in a population that has had limited exposure to the full range of career and stream options available to them. The instrument is working as intended — surfacing profile patterns that require counsellor engagement. The 60.5% of students with strong margin profiles (≥ 10 points) represent the clearest cases: for these students, the leading category can generally be reported with confidence following the five-step algorithm and counsellor review. The 32.5% who triggered at least one Counselling-Required flag are precisely the students who most need structured guidance, and the flag system identifies them efficiently.

8.3 Integrated Use of AAB-Aptitude Assessment Battery and EIP

The AAB-Aptitude Assessment Battery and EIP are designed for integrated use in school career guidance. The AAB-Aptitude Assessment Battery identifies what a student can do — their relative strengths across cognitive and psychomotor aptitude domains — while the EIP maps what a student wants to do — their relative interest in different educational streams. Together, they provide complementary information that

a counsellor can synthesise in discussion with the student.

In practice, the most useful outcomes arise when aptitude and interest profiles converge: a student with high Language–Spatial–Numerical aptitude who also shows a strong Science interest on the EIP has both the ability profile and the interest profile consistent with Science stream. Divergence between the two instruments (high interest in Science but weaker aptitude profile for LSN abilities, for example) does not indicate that one instrument is wrong — it opens a counselling conversation about the student's goals, the demands of the preferred stream, and possible alternative pathways.

The counsellor should always treat both profiles as starting points for structured dialogue, not as definitive prescriptions. Stream recommendations derived from these instruments should be reviewed in discussion with the student, taking account of their academic record, personal goals, and family context.

9. Conclusion

This paper has presented the development and validation of two locally standardized career assessment instruments for Class 10th students in Punjab schools. The AAB-Aptitude Assessment Battery demonstrates satisfactory psychometric properties as a differential aptitude battery. Phase 1 ($N = 521$) established the item base and normative data; Phase 2 ($N = 101$) confirmed battery-level reliability ($\alpha = .7244$, $SB-\rho = .8402$) and revealed a theoretically coherent three-factor structure explaining 86.23% of variance with strong discriminant validity between the three identified aptitude clusters.

The EIP provides a psychometrically sound, ipsatively appropriate interest profile with split-half reliabilities ranging from moderate to very high and strong area discrimination across all five educational areas. The 45.7% odd–even recommendation agreement establishes clearly that the EIP functions as a counselling dialogue tool: it frames and initiates the conversation about educational direction, and the counsellor confirms and refines that direction with the student.

Used together, the AAB-Aptitude Assessment Battery and EIP offer a complementary ability-and-interest picture that supports evidence-based stream selection guidance at one of the most consequential decision points in the Indian secondary education system. Future research should address concurrent validity against stream enrolment and academic outcomes, test–retest reliability, and measurement invariance across gender and school type in order to further establish these instruments' fitness for the population they serve.

References

1. Ackerman, P. L., & Heggstad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, 121(2), 219–245.
2. Aitchison, J. (1986). *The statistical analysis of compositional data*. Chapman and Hall.
3. Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Prentice Hall.
4. Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1984). *Differential Aptitude Tests* (5th ed.). Psychological Corporation.
5. Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, 71(3), 460–502.
6. Brown, A., & Maydeu-Olivares, A. (2012). Fitting a Thurstonian IRT model to forced-choice data using Mplus. *Behavior Research Methods*, 44(4), 1135–1147.

7. Carroll, J. B. (1993). Human cognitive abilities: A survey of factor-analytic studies. Cambridge University Press.
8. Clemans, W. V. (1966). An analytical and empirical examination of some properties of ipsative measures. Psychometric Monographs, No. 14.
9. Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Lawrence Erlbaum.
10. Ebel, R. L., & Frisbie, D. A. (1991). Essentials of educational measurement (5th ed.). Prentice Hall.
11. George, D., & Mallery, P. (2003). SPSS for Windows step by step: A simple guide and reference (4th ed.). Allyn & Bacon.
12. Hansen, J. C., & Campbell, D. P. (1985). Manual for the SVIB-SCII (4th ed.). Consulting Psychologists Press.
13. Holland, J. L. (1985). Making vocational choices: A theory of vocational personalities and work environments (2nd ed.). Prentice Hall.
14. Holland, J. L. (1997). Making vocational choices (3rd ed.). Psychological Assessment Resources.
15. Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. Psychological Bulletin, 96(1), 72–98.
16. Kaiser, H. F. (1974). An index of factorial simplicity. Psychometrika, 39(1), 31–36.
17. Kline, P. (2000). The handbook of psychological testing (2nd ed.). Routledge.
18. Kulshrestha, S. P. (1989). General Aptitude Test Battery — Indian adaptation. National Psychological Corporation, Agra.
19. MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. Psychological Methods, 4(1), 84–99.
20. National Psychological Corporation, Agra-DATB, EIR, PTI and other Psychometric assessment tools and manuals.
21. Rust, J., & Golombok, S. (2009). Modern psychometrics: The science of psychological assessment (3rd ed.). Routledge.
22. Swanson, J. L., & Fouad, N. A. (2015). Career theory and practice: Learning through case studies (3rd ed.). SAGE.
23. Tinsley, H. E. A., & Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. Journal of Counseling Psychology, 34(4), 414–424.
24. U.S. Department of Labor. (1970). Manual for the USES General Aptitude Test Battery. U.S. Printing Government Office.