

# PJRF

## *Professional Judgment Rating Form*

Novice / Internship Level

A Measure of Critical Thinking Abilities and  
Habits of Mind



© 1998, 2006, California Academic Press, San Jose, CA

**Permission is hereby granted for individuals, non-profit organizations, government agencies and educational institutions to download, copy, and distribute the PJRF and this User Manual for non-commercial educational purposes.**

ISBN 978-1-891557-99-6

**Published by Insight Assessment**

Phone: (650) 697-5628 [www.insightassessment.com](http://www.insightassessment.com)

**PROFESSIONAL JUDGMENT RATING FORM:  
NOVICE / INTERNSHIP LEVEL  
CRITICAL THINKING ABILITIES AND HABITS OF MIND**

**User=s Manual**

This User's Manual supports **Professional Judgment Rating Form: Novice / Internship Level (PJRF)**. *Part 1B Theory* focuses on the theoretical construct grounding critical thinking skills and dispositions. *Part 2 B Application* presents basic instructions and information on how to use, score, and interpret the **PJRF** for professional development, individual assessment, program evaluation, and research purposes.

Technical consultation regarding administration of the **PJRF** is available through:

**Insight Assessment, 1735 N 1<sup>st</sup> Street, Suite 306, San Jose, CA 95112**

**Phone: (650) 697-5628.**

**Fax (650) 692-0141**

[www.insightassessment.com](http://www.insightassessment.com)

# TABLE OF CONTENTS

## PART 1 - THEORY

CRITICAL THINKING AND PROFESSIONAL JUDGMENT	1
RELIABILITY AND VALIDITY OF THE <b>PJRF</b> – QUICK NOTES	2
THE DISPOSITION TOWARD CRITICAL THINKING IN PROFESSIONAL PRACTICE	3
THE SEVEN HABITS OF MIND THAT DISPOSE ONE TOWARD CRITICAL THINKING	4
WORKPLACE APPLICATION OF THE SEVEN CRITICAL THINKING HABITS OF MIND	5
CRITICAL THINKING SKILLS IN PROFESSIONAL PRACTICE	6

## PART 2 - APPLICATION

USING THE <b>PJRF</b>	14
SCORING AND INTERPRETING THE <b>PJRF</b>	15
INTERPRETING THE 20 <b>PJRF</b> DESCRIPTIVE STATEMENTS	17
THE POSITIVE ITEMS	17
THE NEGATIVE ITEMS	18
SCORING THE <b>PJRF</b>	21
INTER-RATER RELIABILITY AND THE <b>PJRF</b>	22
CALCULATING AN INTER-RATER RELIABILITY COEFFICIENT	22
THE PROFESSIONAL JUDGMENT CONSTRUCT	25
USES OF THE <b>PJRF</b>	28

REFERENCES	31
------------	----

<b>SAMPLE PROFESSIONAL JUDGMENT RATING</b>	
FORM	33

## PART 1 B THEORY

# CRITICAL THINKING AND PROFESSIONAL JUDGMENT

Professionals are expected to exercise sound, unbiased judgment in interpreting and analyzing information, determining the nature of problems, identifying and evaluating alternative courses of action, making decisions, and, throughout, monitoring the process and impact of their problem solving activity so as to amend, revise, correct, or alter their decisions, or any element that led up to those decisions, as deemed necessary. Judgment in professional practice, correctly exercised, is a reflective, self-corrective, purposeful thinking process which requires the professional to take into account content knowledge, context, evidence, methods, conceptualizations, and a variety of criteria and standards of adequacy. Professional judgment is what educators have called *critical thinking*, but exercised in a practical, professional setting.<sup>1</sup>

Using critical thinking as the foundational concept, as a working definition *professional judgment* can be characterized as a goal-oriented decision-making or problem-solving process carried out in the interest of one's client wherein one gives reasoned consideration to relevant information, criteria, methods, context, principles, policies, and resources.<sup>2</sup> Independent surveys of employers, policy-makers, and educators show overwhelming accord on the importance and centrality of critical thinking to workplace success and the endorsement of the specific critical thinking skills and habits of mind addressed by the *Professional Judgment Rating Form*.<sup>3</sup>

The complexity of professional judgment has direct implications for the education of novice and more advanced practitioners. The exercise of core critical thinking skills, such as analysis, interpretation, inference, evaluation, explanation, and self-correction, is essential to the work of the millions who are program directors, administrators, supervisors, managers, military officers, health care providers, customer service representatives, law enforcement officials, educators, engineers, journalists, ministers, athletes, business agents, and entertainers. Poor thinking can easily be as costly as inexperience or inadequate knowledge of the professional field. But not every lapse of thinking is a failure of skills. Arguably indifference, mental laziness, or inattention lead to more mistakes than do inadequate analyses or unwarranted inferences. In addition to being capable of sound judgment, practicing professionals must be alert to situations requiring the use of their thinking skills and inclined to use those skills in those situations. The two equally required, although conceptually distinctive, dimensions of professional judgment are the *ability to think* and the *willingness to think*.

The assessment of professional judgment requires attention to both the ability to make sound professional judgments and the inclination to do so. The expression "willing and able" is apt. Skill without disposition is a non-starter. Disposition without skill is a poor finisher. Development as a professional requires sharpening the skills and nurturing the dispositions. And this means that we must be able to recognize their manifestation in a novice professional's problem-solving behavior and attitude toward using thinking, as opposed to some other strategy, to engage and resolve problems.<sup>4, 5</sup> Rating made using the **PJRF** focus the novice as well as the mentor who performs the evaluation on the attitudinal and the skill dimensions of thoughtful problem-solving. The sections that follow describe the attitudinal and skill elements so vital to good thinking in professional practice contexts.

The **PJRF** is intended for use by experienced professionals in judging the extent to which novices approach problems with critical thinking. The **PJRF** complements **The California Critical Thinking Skills Test** which is an objective measure of critical thinking skills, and **The California Critical Thinking Disposition Inventory** which can provide a profile of a person's critical thinking inclinations. For information about these other tools, contact the California Academic Press or visit our website at [www.insightassessment.com](http://www.insightassessment.com)

### Ethics and Performance Testing

Use, administration, scoring, and interpretation of the PJRF and other educational assessment tools published by the California Academic Press is the sole responsibility of the purchaser and user. The California Academic Press strongly recommends that persons using the PJRF and other educational testing tools should observe ethical, academic, and professional standards in the use, administration, scoring, and interpretation of those instruments.

Many professional associations have published guidelines regarding ethical practices in educational testing. One might consult "Principles of Good Practice for Assessing Student Learning," issued in December 1992 under the auspices of the American Association of Higher Education, One Dupont Circle, Suite 360, Washington DC, 20036; "Ethical Standards of the American Educational Research Association," issued in October 1992, AERA, 1230 17th Street, NW, Washington DC 20036; and "Standards for Educational and Psychological Testing," issued in 1985 by the American Psychological Association Inc., also at 1200 17th Street, NW, Washington DC, 20036.

# Reliability and Validity of the PJRF – Quick Notes

## *Validity*

### *Reliability*

The **PJRF** may be used by an individual to make evaluations. In this case the judge(s) should strive for consistency over time and from case to case. If one plans to use this tool in a more formally designed evaluation project, or if multiple judges wish to assess their mutual agreement and consistency, reliability coefficients can be estimated.

For many assessment tools test/retest reliability, alternate form reliability, and internal consistency estimates such as coefficient alpha are reported. Such reliability estimates are not reported here because the reliability of the **PJRF** (and all other expert rating forms) completely depends on the judges using the form.

Reliability for the **PJRF** can be established on a site-to-site basis. Reliability of the **PJRF** depends on judges' ability to be consistent with themselves and the degree to which agreement can be established between judges. The degree of agreement between judges can be estimated and is known as inter-rater reliability. In the Application section of this manual (beginning on page 13), we explain how to train judges to increase the inter-rater reliability between them. We also explain how to calculate an estimate of inter-rater reliability.

All of the prompts found on the PJRF are derived from The American Philosophical Association's Delphi consensus definition of critical thinking described in detail in the sections below. The APA Delphi consensus conceptualization of critical thinking (CT) is an historically important benchmark.<sup>2</sup> It is an expression of expert consensus articulated without the constraints of accreditation or legislation, and based on the participation of 46 leading theorists, teachers, and CT assessment specialists from several disciplines. Further confirmation of the validity of this construct derives from the 1993-1994 national survey and replication study conducted by the National Center for Higher Education Teaching, Learning and Assessment at The Pennsylvania State University.<sup>3</sup>

The **PJRF**, the **CCTST**, the **CCTDI**, and the [Holistic Critical Thinking Scoring Rubric](#), are the first CT assessment instruments to derive their construct validity from the APA Delphi Report conceptualization. As such, they represent a significant advance over previously published CT instrumentation.

## The Disposition Toward Critical Thinking in Professional Practice

The disposition toward critical thinking is the **consistent internal motivation** to engage problems and make decisions by using thinking.<sup>6</sup> National surveys of employers, policy-makers, as well as educators, repeatedly find the consensus that the dispositional as well as the skills dimension of critical thinking should be considered an essential outcome of a college education.<sup>2,3</sup> For professional preparation at the collegiate level, the teaching of thinking must include, to the extent possible, the sharpening of students' cognitive capabilities and the nurturing of those habits of mind which alert students to opportunities to use thinking to resolve problems and incline students toward doing so.

## The Seven Habits of Mind that Dispose One Toward Critical Thinking

Whether experts or novices, some approach problems confident in their own ability to reason them through, others mistrust themselves as decision-makers, thinkers, or problem-solvers. Some people are open-minded, others are intolerant. Some seek to approach problems in diligent, focused, and systematic ways; others tend to be scattered, disorganized, and easily

distracted. Some seek for evidence and reasons as they consider what to do; others eschew data and principled approaches preferring rather to decide on the basis of impulse, whim, fashion, pressure, or caprice. Some have high levels of professional integrity, others seem unable to step past personal biases, fears, self-interest, or preconceptions. Some see the complexity and subtlety of problems and note multiple possible resolutions; others see things in starker, more dualistic terms, such as good or bad, right or wrong, true or false, black or white. Some people are curious as to the workings of things, they wish to know much more about the problem than simply how to resolve it; others are more than content just to know what to do, and prefer not deal with the why or the what if.

These attitudes, values, and inclinations are dimensions of one's personality which relate directly to how successful we can expect different people to be in the application of their reasoning skills to problem framing and problem resolution. The characteristics which describe the ideal critical thinker have been determined more precisely through two independent studies using social psychological research methods.<sup>1,7</sup> Seven distinctive elements emerged within that description when statistical techniques of factor analysis were applied. In their positive manifestation, these seven characterological attributes or habits of mind, if you prefer, are named *truth-seeking, open-mindedness, analyticity, systematicity, CT self-confidence, inquisitiveness, and maturity of judgment*.<sup>6,7</sup>

Of course, there are vices which are the negatives of each of the seven virtuous thinking by being *intellectually dishonest, intolerant, inattentive, haphazard, mistrustful of reason, indifferent, and simplistic*. While the virtues are obvious assets to the practicing professional, the vices are perhaps even more obvious liabilities. From the employer=s perspective, what becomes interesting is how persons with different profiles will tend to respond to problem-solving situations in professional practice settings. The success of the individual and of the organization will depend on the appropriate exercise of professional judgment in a given context.<sup>1</sup>

## Workplace Application of the Seven Critical Thinking Habits of Mind

One good way of seeing the significance of these habits of mind is to ask, what are the individual, organizational, and societal costs when judgment is entrusted to professionals who are ambivalent or negative on one or more of those seven?

*Truth-seeking* measures the **intellectual honesty** which gives one the courageous desire for best knowledge in any situation, the inclination to ask challenging questions and to follow the reasons and evidence wherever they lead. How many managers have failed their companies by shying away from the hard questions, discounting important and relevant but unpleasant data, and preferring to hold tight

habits of mind. The antithesis of the ideal would be a person who habitually approached to outmoded ways of doing business or untested assumptions about the marketplace?

*Open-mindedness* measures **tolerance** for new ideas and divergent views. What are the chances of helpful critique or innovation if one is intolerant and closed-minded?

*Analyticity* measures alertness to problem situations and potential difficulties which means **being alert to the need to intervene by the use of reason and evidence** to solve problems. A health care professional, an attorney, a teacher, a manager, an engineer, or a policy-maker who is not inclined toward analyticity will likely fail to anticipate significant consequences and, thereby, increase the risk of malpractice and negligence.

*Systematicity* measures the inclination to be **organized**, including to be focused, diligent, and persevering. How much business will be lost by a customer service representative whose approach to clients is disorganized, unfocused, sloppy, and half-hearted?

*Inquisitiveness* measures intellectual curiosity and the intention to learn things even if their immediate application is not apparent. Workers who are indifferent or disdainful of learning more than the minimum necessary to get through the day=s tasks should not expect frequent and speedy promotions.

decision making. Those who see everything as starkly good or bad, black or white, right or wrong, are unlikely to be sophisticated learners or good candidates for positions of increasing responsibility. They are apt to make decisions too hastily or too slowly, to be unwilling to reconsider, to be dogmatic and dualistic, if not outright simplistic, in their approach to problem solving, and to lack sensitivity to the nuances of circumstances and subtleties of context.

*Critical thinking self-confidence* measures **the trust one places in one's own reasoning and one's ability to guide others to make rational decisions.** A career involving any form of thoughtful decision making or problem solving using mediation is not encouraged for persons weak in reasoning self-confidence. But, if their reasoning self-confidence is well founded, that is, if they have the strong critical thinking skills as well as the disposition to use them, they can become quite successful in a wide variety of executive, managerial, client service, and professional occupations.

The [California Critical Thinking Disposition Inventory](#) provides a scientific measure of these critical thinking habits of mind. This tool is now available on line and in paper-and-pencil form. The CCTDI has enjoyed great popularity and is now available in many translations. Consult the Insight Assessment website for details.

*Cognitive maturity* measures **judiciousness** which inclines one to see the complexity in problems and to desire prudent

## **Critical Thinking Skills in Professional Practice**

An international consensus of experts in critical thinking identified six central skills as forming the core of critical thinking, no matter in what professional area, academic field, or everyday life situation it is used. These skills are familiar to good thinkers everywhere on the planet. They are the skills of **interpretation, analysis, inference, evaluation, explanation and self-correction.** In theory we can speak of them as different skills, in practice they function together, like the parts of a finely tuned engine, to drive the mind toward understanding a problem, finding and evaluating alternatives, deciding possible consequences of various choices, looking to results to see what might be done better, and monitoring the entire process to be sure that one is not making important mistakes. <sup>2</sup>

When mentoring a new employee or in evaluating an individual's performance, one can put more emphasis on one skill than on another. Yet we must always remember that in actual practice when a person uses any one of them the person is actually using all of them to some extent.

Each of the six primary critical thinking skills can be defined and broken down into its sub-skills. The examples under each are useful in clarifying what is meant by the different skills so that people can parts of the thinking situation. Again, like the analogy to the engine, however, even if all the parts are in good order, it is not until the engine is assembled and tuned that it can actually generate results. So too with

**Interpretation-- Categorization:**

thinking, the specific skills or sub-skills by themselves are of little use unless they are assembled into the overall process of making a judgment about what to do or what to believe in a given professional context.

**INTERPRETATION:** To comprehend and express the meaning or significance of a wide variety of experiences, situations, data, events, judgments, conventions, beliefs, rules, procedures or criteria.

The sub-skills of **Interpretation** are: **Categorization, Decoding Significance, Clarifying Meaning**

- § to apprehend or appropriately formulate categories, distinctions, or frameworks for understanding, describing or characterizing information.
- § to describe experiences, situations, beliefs, events, etc. so that they take on comprehensible meanings in terms of appropriate categorizations, distinctions, or frameworks.

For example: to recognize a client=s problem and define its character without prejudice; to determine a useful way of sorting and sub-

communicate effectively about one another=s thinking. This is the key to the effective use of a tool like the **PJRF**, for it allows the novice and the mentor to focus attention on specific classifying relevant information about a case; to make an understandable report of what one experienced in a given situation, for example an interview; to classify data, findings, or opinions using the classification system provided by one=s employer.

**Interpretation -- Decoding Significance:**

§ to detect, attend to, and describe the informational content, emotional shape, directive functions, intentions, motives, purposes, social significance, values, views, rules, procedures, criteria, or inferential relationships expressed in convention-based communication systems, such as in language, social behaviors, drawings, numbers, graphs, tables, charts, signs and symbols.

For example: to detect and describe a person's real intention in asking a given question or raising a given issue; to appreciate the significance of a particular facial expression or gesture used in a given social situation; to discern the use of irony or rhetorical questions in a conversation or in an argument; to interpret the data displayed on a spreadsheet or in a diagram or using the technology or special instruments appropriate to a given professional field of relevance to the problem being discussed.

**Interpretation -- Clarifying Meaning:**

§ to paraphrase or make explicit, through language, diagrams, definitions, description, analogy or figurative

expression, the contextual, conventional or intended meanings of words, ideas, concepts, statements, behaviors, drawings, numbers, signs, charts, graphs, symbols, rules, events or ceremonies.

§ to use description, definitions, analogy, pictures, diagrams, or figurative that person's intended meanings; to find an example which helps explain something to someone; to develop a distinction which makes clear an important difference or removes a troublesome ambiguity; to diagram the physical or other relationships between objects under discussion.

**ANALYSIS:** To identify the relationships among the elements in a larger whole; in this case the reasoning relationships between ideas, statements, questions, concepts, descriptions, diagrams, data, or other forms of representation intended to express a person=s or group=s beliefs, judgments, experiences, views, concerns, interests, reasons, information, or opinions.

§ to identify issues, determine their component parts, and to express the relationships of those parts to each other and to the whole.

The sub-skills of **Analysis** are: **Examining Ideas, Identifying Arguments, and Analyzing Arguments**

**Analysis -- Examining Ideas:**

§ to determine the role various expressions play, or are intended to play, in the context of a conversation, argument, text, reasoning situation, debated, or effort at persuasion.

§ to define one=s terms.

expression to remove confusion, vagueness, or ambiguity, or to design a reasonable procedure for so doing.

For example: to restate what a person said using different words or expressions while preserving

§ to compare or contrast ideas, data, concepts, or statements.

For example: to identify a phrase intended to trigger a sympathetic emotional response that might induce an audience to agree with an opinion; to examine two or more closely related proposals regarding a given problem and to determine their points of similarity and divergence; given a complicated assignment, to determine how it might be broken up into smaller, more manageable tasks; as a step in mediation, to identify clients= several possibly divergent or convergent interests in a given situation and to express the relationship between those interests.

**Analysis-- Detecting Arguments:**

§ given an expression of a point of view, or a group of statements, descriptions, questions, or diagrams or other graphic representations, to determine whether or not that expression or body of information is intended to express a reason or reasons in support of, or contesting, some claim, opinion, or point of view.

For example, given a paragraph from a memo, determine whether a standard reading of that paragraph, in the context of how and where it is published, would suggest that it presents a claim as well as a reason or reasons in support of that claim; given a passage from a report or commentary from an analyst, determine if the author of that passage intended it as an expression of reasons for or against a given claim, or opinion, or proposed course of action;

given a commercial announcement, identify any claims being advanced along with the reasons presented in their support; given the questions or concerns of clients, determine whether the clients intend to be stating their reasons for or against accepting a particular course of action.

**Analysis: Analyzing Arguments:**

§ given the expression of a reason or reasons intended to support or contest for those premises and reasons intended as supporting the main conclusion, (d) additional unexpressed elements of that reasoning, such as intermediary conclusions, unstated assumptions or presuppositions, (e) the overall structure For example: given a document expressing a legal opinion, or a discussion paper on a of the argument or intended chain of reasoning, and (f) any items contained in the body of expressions being examined which are not intended to be taken as part of the reasoning being expressed or its intended background. controversial issue, to identify the author's chief claim, the reasons and premises the author advances on behalf of that claim, the background information the author used to support those reasons or premises, and crucial assumptions implicit in the author's reasoning; given a series of background papers on a topic of professional concern, find within them several reasons or chains of reasons in support of or in opposition to responding in certain way, develop a graphic representation to illustrate what the reasons are and how they converge or diverge around doing accepting a certain conclusion about what to think or what to do.

**INFERENCE:** To identify and secure elements needed to draw reasonable conclusions; to form conjectures, find alternatives, and generate hypotheses; to consider relevant information and to educe the consequences flowing from data, statements, principles, evidence, judgments, beliefs, opinions, ideas, graphics, concepts,

some claim, opinion or point of view, to identify and differentiate: (a) the intended main conclusion, (b) the premises and reasons advanced in support of the main conclusion, (c) further premises and reasons advanced as backup or support

descriptions, questions, or other forms of representation.

The sub-skills of **Inference** are :**Querying Evidence, Conjecturing Alternatives, and Drawing Conclusions**

**Inference -- Querying Evidence:**

- § in particular, to recognize assumptions, reasons, or statements that require support and then to formulate an effective strategy for seeking and gathering the information that might supply that support.
- § in general, to judge what information or understandings might be relevant to deciding the acceptability, plausibility or relative merits of a given alternative, question, claim, issue, theory, hypothesis, or statement is required, and to determine plausible and effective investigatory strategies for acquiring that information.

For example: when attempting to develop a persuasive presentation in support of one's goals, judge what background information it would be useful to have and to develop a plan which will yield a clear answer as to whether or not such information is available; after judging that certain missing information would be germane in determining if one opinion is more or less reasonable than a competing opinion, to plan

a search which will reveal if the information crucial to deciding that question is available.

**Inference -- Conjecturing Alternatives:**

- § to formulate plausible alternatives for resolving a problem, to postulate a series of suppositions regarding a question, to project alternative hypotheses regarding an why something happened, to develop a variety of different plans to achieve some goal.
- § to draw out the presuppositions and to project the range of possible consequences, both positive and negative, of alternative decisions, positions, policies, theories, or beliefs.

For example: given a problem with technical, budgetary, or ethical ramifications, to develop a set of options for addressing and resolving that problem; given a set of organizational priorities with which one may or may not agree, to project the difficulties and as well as the benefits that are likely to result if those priorities are adopted in decision making.

**Inference -- Drawing Conclusions:**

- § to apply appropriate modes or strategies of reasoning to determining what
- § position, opinion, or point of view one should take on a given matter or issue.
- § given a set of statements, descriptions, questions, data, assumptions, or other forms of representation, to educe, with the proper level of logical strength, the possible, probable, and necessary consequences that they support, warrant, imply or entail.
- § to employ successfully various sub-species of reasoning, as for example to reason using appropriate statistics, analogies, mathematical formulae, precedents, etc. as appropriate to the context, information, and issues at hand.

to determine which of several possible conclusions is most strongly warranted or supported by the evidence at hand, or which should be rejected or regarded as less plausible by the information given. For example: to carry out experiments and to apply appropriate statistical inference techniques in order to confirm or disconfirm a testable hypothesis; given a controversial issue, to examine informed opinions, consider various opposing views and the reasons advanced for or against them, gather relevant information, and formulate a considered opinion regarding that issue; to deduce a theorem from axioms using prescribed rules of inference; in determining a course of action for a given case at hand, to apply relevant policies and precedents of the organization.

**EVALUATION:** To assess the credibility of statements or other representations that are accounts or descriptions of a person's perception, experience, situation, judgment, belief, or opinion; and to assess the strength of the reasoning relationships among statements, graphics, data, descriptions, questions or other forms of representation.

The sub-skills of **Evaluation** are:  
**Assessing Claims, and Assessing Arguments**

**Evaluation -- Assessing Claims:**

- § to recognize the factors relevant to assessing the degree of credibility to ascribe to a person or to a given source of information or opinion.
- § to assess the contextual relevance of questions, information, principles, rules or procedural directions.
- § to assess the acceptability or the level of confidence to place in the probability or truth of any given assertion,

representation of a person=s experience, situation, judgment, belief or opinion.

- For example: to recognize the factors which make a given person a more or less credible witness regarding a given event or credible authority on a given topic; to determine if a given analogy to past decisions is or is not applicable to deciding what to do in a new and somewhat similar situation; to determine if a given claim or idea that is based upon those data, beliefs, or
- § to determine the extent to which possible additional information might strengthen or weaken an argument. statements.
  - § to anticipate or to raise questions or objections, and
  - § to assess whether these point to significant weakness in the argument being evaluated. to determine whether an argument relies on false or doubtful assumptions or presuppositions and then to determine how crucially these affect its strength.
  - § to judge between reasonable and fallacious inferences.
  - § to judge the probative strength of an argument's premises and assumptions with a view toward determining the acceptability of the argument.
  - § to determine and judge the probative strength of an argument's intended or unintended consequences with a view toward judging the acceptability of the argument;

For example: given a specific claim someone is making, to judge if that claim follows either with certainty or with a high level of probability from the reasons the person advances on its behalf; to check for identifiable formal and informal fallacies; given an objection to an argument to evaluate the reasonableness or logical force of that objection; to judge the quality and applicability of analogical arguments; to judge the

being talked about by one=s co-workers is likely to be true or likely to be false based on what one knows or can reasonably find out.

#### **Evaluation -- Assessing Arguments:**

- § to judge whether the assumed acceptability of a set of statements, beliefs, or data justifies one's then accepting as necessarily or very probably true the expressed conclusion or claim
- logical strength of arguments based on hypothetical situations or causal reasoning; to judge if a given argument is relevant or applicable or has implications for the situation at hand; to determine how possible new data might confirm or disconfirm a given opinion.

**EXPLANATION:** To state the results of one's reasoning; to justify that reasoning in terms of the evidence, methods, concepts, standards, and contextual considerations upon which one's results were based; and to present one's reasoning in the form of cogent arguments.

#### The sub-skills of **Explanation** are: **Stating Results, Justifying Procedures, and Presenting Arguments**

#### **Explanation -- Stating Results:**

- § to produce accurate statements, descriptions or representations of the results of one's reasoning activities so one may interpret, analyze, evaluate, infer from, or monitor those results.

For example: to state one's reasons for holding a given view; to write down for future use one's current thinking about an important or complex matter; to state what one has learned as a result of an investigation, reflection, experience, or experiment of some kind; to convey one's analysis and judgment regarding the quality a work of art in light of appropriate standards; to

state one's considered opinion on a matter of practical urgency, giving one's reasons for that opinion.

**Explanation -- Justifying Procedures:**

§ to present the evidence, concepts, methods, criteria, and contextual considerations which one used in forming one's interpretations, analyses, evaluation or inferences, so that one might accurately record, evaluate, describe or justify those processes to one's self or to others, or so as to remedy perceived deficiencies in the general way one executes those processes.

For example: to keep a log of the steps followed in working through a long or difficult problem, new experience, decision-making process, or scientific procedure; to explain one's choice of a particular graphic representation or spreadsheet array of data, or the use of a given statistical test for purposes of data analysis; to state the standards one should use in evaluating a claim, or client's expression of concern, or set of conflicting priorities; to explain how one understands a key idea or policy; to show that the prerequisites for the use of a given technology have been satisfied; to report the strategy one used in attempting to make a decision in a reasonable way; to design and explain a graphic display which represents the quantitative or spatial information used as evidence in making a decision.

**Explanation -- Presenting Arguments:**

§ to give reasons for accepting some claim.  
§ to meet objections to the method, conceptualizations, evidence, criteria or contextual appropriateness of one's inferences, analyses, interpretations, or evaluative judgments.

For example: to write a position paper; to anticipate and to respond objectively to reasonable criticisms one might expect to be raised against one's political views; to reason in collaboration with others and to assist in expressing the methods, ideas, standards, evidence, and contextual factors that the group is and should be using to reach its conclusions; to identify and to express evidence and counter-evidence intended to refine one's own or another person's thinking on a matter of concern.

**SELF-REGULATION:** Self-consciously to monitor one's cognitive activities, the elements used in those activities, and the results educed, particularly by applying skills in analysis and evaluation to one's own inferential judgments with a view toward questioning, confirming, validating, or correcting either one's reasoning or one's results.

The sub-skills of **Self-Regulation** are: **Self-Examination, and Self-Correction.**

§ **Self-Regulation -- Self-Examination:** to reflect on one's own reasoning and verify both the results produced and the correct application and execution of the cognitive skills involved.

§ to make an objective and thoughtful meta-cognitive self-assessment of one's opinions and reasons for holding them.

§ to judge the extent to which one's thinking is influenced by deficiencies in one's knowledge, or by stereotypes, prejudices, emotions or any other factors which constrain one's objectivity or rationality.

§ to reflect on one's motivations, values, attitudes and interests with a view toward determining that one has endeavored to be unbiased, fair-minded, thorough, objective, respectful of the truth,

reasonable, and rational in coming to one's analyses, interpretations, evaluations, inferences, or expressions. For example: to examine one's views on a controversial issue with sensitivity to the possible influences of one's personal bias or self-interest; to review one's methods or calculations with a view to detecting mistaken applications or inadvertent errors; to reread sources to assure that one has not overlooked important information; to identify and review the acceptability of the facts, opinions or assumptions one relied on in coming to a given point of view; to identify and review one's reasons and reasoning processes in coming to a given conclusion.

**Self-Regulation -- Self-Correction:**

§ where self-examination reveals errors or deficiencies, to design reasonable procedures to remedy or correct, if possible, those mistakes and their causes.

For example: given a methodological mistake or factual deficiency in one's work, to revise that work so as to correct the problem and then to determine if the revisions warrant changes in any position, findings, or opinions based thereon.

[Health Sciences Reasoning Test](#), and [The Law Student Reasoning Profile](#), among others offered by Insight Assessment provide objective measures of these critical thinking skills. These tools return scores on one's overall critical thinking skill, as well as subscale scores on deductive-reasoning, inductive-reasoning, analysis-interpretation, inference, and evaluation-explanation. Visit the Insight Assessment website for a complete listing of authorized translations for these critical thinking skills tests. These tools are available in both on-line e-testing and paper-and-pencil formats.

Designed for use with different audiences and educational levels, the [California Critical Thinking Skills Test](#), the [Test of Everyday Reasoning](#), and

## PART 2 B APPLICATION

### USING THE PROFESSIONAL JUDGMENT RATING FORM

The critical thinking skills and habits of mind described above form one of the three pillars of sound professional judgment. The other two are background knowledge of the professional field and reflective experience. Development into a successful practitioner cannot result if any of the three are missing. The *Professional Judgment Rating Form* incorporates both the skills and the habits of mind dimensions of critical thinking, one of those three essential pillars. The next sections of this Manual describe that tool and how it should be used. A sensitive mentor will combine the use of the **PJRF** with guidance on how to learn more of the information needed to be a successful practitioner and how to gain the most out of one's reflections on one's initial and formative experiences as a novice professional.

The assessment of professional judgment requires attention to both the ability to make sound professional judgments and the inclination to do so. As we said earlier, the expression "willing and able" is most apt. The possession of thinking skills without the inclination to use them means that a person is likely to approach problems using some other, probably less reliable and less rational, strategy. Yet if one is disposed toward using thinking, without the skill to think well, means that the person will arrive at a flawed judgment or an inadequate resolution to a problem. In both cases, there is little chance of sustained success as a practicing professional. These theoretical considerations have been reinforced by independent researchers in many different fields.<sup>8-12</sup>

Development as a professional requires sharpening the critical thinking skills of analysis, inference, interpretation, evaluation, explanation, and self-correction as described in the theory section of this Manual. Development also implies nurturing critical thinking habits of mind such as truth-seeking, open-mindedness, analyticity, systematicity, confidence in one's own reasoning, inquisitiveness, and maturity of judgment as those were described earlier. As one begins practice as a novice professional or intern, the things one does and says, how one responds to problem situations, and how one makes decisions all tend to manifest the quality of one's critical thinking habits of mind and the level of one's critical thinking skills.

Evidence of positive habits of mind with respect to making professional judgments is not that difficult to observe, if one is looking. In films, novels, and real life it is not uncommon to see professionals deeply engaged in the process of resolving ill structured problems; objectively

seeking reasons and evidence to support decisions, claims, accusations; endeavoring systematically and analytically to identify problems, interpret situations, and anticipate consequences; inquiring about new knowledge, following up on possible sources of new information; reconsidering past decisions in the light of new information; and avoiding hasty judgments but moving toward necessary decisions in a timely way. In many novels, films, and real life contexts we have seen the potentially disastrous consequences of not exercising good judgment.

To facilitate and make operational the measurement of professional judgment in the academic settings and the workplace requires efficient instruments which can be implemented by practicing professionals as part of employee development. Useful tools can be built if one draws on a sound theoretical understanding of critical thinking and the pragmatic sensitivity to the weak or strong thinking which is so evident in actual professional practice. For professional workplace settings we suggest the newly developed *Professional Judgment Rating Form: Novice / Internship Level*.

## **Scoring and Interpreting The Professional Judgment Rating Form**

The ***Professional Judgment Rating Form -- Novice / Internship Level*** consists of two groups of ten descriptors. The first group of ten relate *primarily* to the attitudinal (or habits of mind) dimension of critical thinking. The second group of ten relate *primarily* to critical thinking skills. If one is displaying strong critical thinking skills, one must be at least somewhat disposed toward critical thinking or the skills would not be displayed at all. When using the **PJRF** it is not necessary for the rater to attempt to separate the skills from the dispositions.

The mentor, supervisor, or person doing the rating should know the individual being evaluated well enough to respond to at least 17 of the 20 descriptors with confidence based on personal knowledge and observation. If one has not personally observed the individual in a workplace setting, for purposes of the validity of the evaluation one is strongly advised not to undertake a formal evaluation of the person by using this measurement tool.

A mentor, an experienced professional, or an internship supervisor evaluates the novice by responding **yes** or **no** to each of the 20 descriptors. The rated individual is then awarded one point for each desirable response the rater provides. The desirable responses, as explained below, are in some cases **Yes** and in others **No**.

The scoring guidelines and suggested interpretations are a baseline for rating a person using the **PJRF**. The resulting scores may be compared, if inter-rater reliability has been established, so that one may track about the same individual over time or so that one may compare different individuals. **PJRF** scores provide a basis for conversation about the manifestation of critical thinking skills and dispositions in a novice=s professional judgment.

If a single supervisor makes ratings over time and strives for consistency, comparisons between ratings can assess changes as a result of staff development. Two or more scores on the PJRF made by two or more raters should not be compared if inter-rater reliability has not been established. Any observed differences may only reflect differences in the raters. Establishing inter-rater reliability is comparable to calibrating one=s instrumentation prior to using it to make important measurements. One seeks to assure that if more than one person is using the tool to examine the same thing, the results obtained are consistent.

## Interpreting the 20 PJRF Descriptive Statements

The successful use of the **PJRF** depends on the fairness, objectivity, and personal knowledge that the rater has of the ways in which the novice approaches and carries out professional decision making and problem solving in the workplace setting. The interpretation and scoring of items described below is a generalized expression of novice judgment.

### The Positive Items: 1, 2, 8, 9, and 10

1. Willing to engage challenging problems in the workplace.

1. Seeks clarification of the problem and the terminology used to discuss it.

**Item #1** in both sections reflects positive attributes. We would want novice professionals to be willing to take on challenging problems, as opposed to avoiding them or as opposed to only being willing to take on easy problems. This shows some self-confidence, obviously; but, by itself, is not enough information. For it does not tell us whether the person is taking on too much too soon. The second item #1 suggests prudence and some skill in defining the problem before trying to solve it. Novices soon discover, if they are going to make any progress at all, that their way of seeing a

problem is often considerably less sophisticated than the way experts describe the problem. Hence evidence that the novice is working to develop skill at problem definition is an important first step.

2. Systematic in approaching problems and available solutions.

2. Seeks reasons and evidence to support offered assertions and evaluations.

The first **item #2** indicates systematicity and concern for focus and organization in problem solving. The second item #2 aims that focus at finding reasons and evidence, the most crucial elements in good thinking. Here the reasons and evidence are directed at support for assertions and evaluations, which are something less than full problem solutions. Again, these are expectations of novices.

8. Willing to reconsider decisions in light of new information.

8. Integrates new information and adjusts direction to resolve a difficult problem.

**Item #8** address how the reasons and evidence are used. A level of open-mindedness is identified as a good thing in the first item #8. Skill at integrating new information and rethinking one=s preliminary

conclusions are positive features addressed in the second item #8.

9. Mindful of relevant considerations when addressing a dilemma.

9. Anticipates consequences likely to occur as a result of decisions.

Attention to relevant considerations is the cornerstone of strength in both inference and explanation, and is addressed in the first item #9. Skill at using these considerations to actually draw out consequences, which is vital if good evaluations are to ensue, is the target of the second **item #9**.

10. Likely to be the person to identify a problem that requires attention.

10. Anticipates potential difficulties and suggests possible responses before problems arise.

Analyticity, or being alert for opportunities to use reason, is the main focus of the first **item #10**; actual skill at being analytical and particularly at correctly anticipating problems is the positive target in the second #10.

## The Negative Items: 3, 4, 5, 6, and 7.

There are as many items devoted to identifying shortcomings in the novices critical thinking habits of mind and skills as there are items aimed at the strengths that are being manifested. Items # 3 through # 7 on both lists are those that look at possible weaknesses.

3. Indifferent to problems constraining performance of co-workers.

3. Belittles the views, opinions, suggestions, or perspectives of others.

The first **item #3** addresses the importance of critical thinking conceived of as a group process, not just an individual exercise. Professional decision making happens often when experts consult with one another and come to a mutual consensus about the nature of a problem and the optimal response. By contrast, when a person's style constrains the success of others in the professional setting, that is clearly a negative. If the person is indifferent to that, then we can infer that the person is failing to analyze and interpret the situation properly, or failing to draw the reasonable inference that constraints on the performance of co-workers will have a

negative impact on the overall success of the office or organization. If the first item #3 is much stronger, for it indicates that the person actually contributes to the problem by displaying intolerance.

Both the #4 items and the #5 items suggest failures of thinking skills and habits of mind.

4. Too quick to define problem situations in ways that neglect relevant context.

4. Applies policy or performs tasks without reflecting on quality or impact.

In the **first #4** the problem is going to be defined poorly, even for a novice, and mistakes will follow. The second #4 also will result in problems for it suggests a mindless approach. While we might well expect that a novice would carry out the policy as written, and not expect that the novice would critique the policy or seek a change, we would still expect that even the novice professional should reflect on the impact of the policy or on the quality of the work being produced.

5. Too quick to discard potential alternative solutions for a problem.

5. Focuses more on the concerns of self or co-workers than on the needs of clients.

The first **item #5** shows weakness in thinking skills and poor habits of mind. Other research suggests that truth-seeking is the

seen as a simple indifference or failure to perceive this, the second item #3 most difficult of all the habits of mind to engender.<sup>13,14</sup> Clearly a person who is hasty and fails to give other ideas a fair hearing is going to have problems with truth-seeking, if not with core skills like analysis and evaluation as well. The second item #5 takes this even further for it describe an aspiring professional whose thinking is not directed, as should be the priority, at problem-solving on behalf of the client.

6. Inclined to respond to problems with familiar but inappropriate strategies.

6. Worries more about the origin of an idea than its quality.

The first **item #6**, if the response is positive, indicates a person who does not think about how best to respond and who does not respond to problems by thinking. Using pat answers, familiar routines, unreflective recipes, scripts, or stock procedures when not appropriate, is not the mark of a good thinker, even if the person is a novice. The second #6 is equally a problem, if the response is positive. For if the person is more concerned with the origins or authorship of an idea than with its quality, either the person is unable or unwilling to apply good thinking to evaluate the idea on its own merits.

7. Intolerant of potential solutions outside of existing protocols or procedures.

7. Generally declines to participate actively in problem solving strategy discussions.

The first **item #7** displays hostility toward thinking. A yes here means the person is failing to come to grips with the content and quality of the ideas and, instead, is looking for inconsistencies with established avenues as an automatic excuse to reject them. Perhaps the only thing less auspicious would be behavior that results in a yes response to the second item #7. In such as case the person is stepping away from problem solving and perhaps should be counseled to consider an alternative career.

It is hard to imagine a novice who generally declines to participate and learn from strategy discussions as having anything but the most dismal prospects for success.

## Scoring the PJRF

When you have confidence that you have observed the novice or intern in the workplace setting sufficiently to make a thoughtful and accurate response to the descriptive statements, think about specific situations that apply and mark the **Yes** or the **No** column to the left of each of the 20 descriptive statements. As a general rule, and unless there is some overriding consideration peculiar to a given setting as determined by a consensus of senior professionals at that setting, apply the PJRF to novices by scoring as follows:

**Count** +1 point for each **YES** on all items 1, 2, 8, 9, and 10.  
+1 point for each **NO** on all items 3, 4, 5, 6, and 7.

- Scores 17 - 20:** **Very Strong:** Repeatedly demonstrates the consistent internal motivation and mental ability to make professional judgments in the workplace. Makes judgments that are mindful of relevant considerations, contexts, methods and criteria. Has a style that reflect a judicious, open-minded and honest approach to judgments in the workplace.
- Scores 13 - 16:** **Positive:** Adequately demonstrates the ability and habits of mind to make professional judgments in the workplace. Generally fulfills the demands for professional judgment of the quality required of individuals holding this position.
- Scores 8-12:** **Marginal / Ambiguous:** Inconsistently demonstrates the ability and motivation for making professional judgments in the workplace. At times appears to lack the motivation or to lack the ability to make judgments of the quality required by individuals holding this position.
- Scores 4 - 7:** **Negative:** Demonstrates the lack of mental ability or personal motivation for making professional judgments in the workplace. Appears indifferent, resistant, closed minded, disorganized or biased. Most often appears to be unable or unwilling to make judgments of the quality required by individuals holding this position.
- Scores 0 - 3:** **Very Poor:** Constant demonstration of lack of thinking skills and a motivation not to make professional judgments in the workplace; has an aversion to honest engagement of reasons and evidence; makes thoughtless, unreflective, self-serving, or uninformed judgments. Has a style which reflects imprudence, intolerance, disorganization, and immaturity of judgment.

## Inter-Rater Reliability and the PJRF

Inter-rater reliability is an estimate of the degree of agreement between scores given by two or more judges on the same test<sup>20</sup>. The same considerations that lead us to be concerned that our scales are balanced and that we know how to use scientific tools properly are those that lead us to appreciate the importance of establishing inter-rater reliability before we begin to gather data and make decisions using rating forms and performance rubrics. People need to be sure that they are viewing similar things in similar ways. Otherwise, what one person thinks of as evidence on behalf of one judgment, the other person could think of as evidence for the opposite. The sports world uses trained referees and umpires who have practiced the application of rules to simulated situations to assure inter-rater reliability when the real games are being played. Judges of difficult to evaluate phenomena, like spring-board diving, figure skating, and piano virtuosity are persons whose judgment in such matters has been refined by a knowledge of the relevant standards, reference to examples of all levels of success and failure, and by practice. They learn to see the quality of a performance the way other trained judges see it. In this section we will discuss how to calibrate judges and establish a high degree of inter-rater reliability as well as describe

how to mathematically compute inter-rater reliability.

### Calculating an Inter-Rater Reliability Coefficient

One way to calculate inter-rater reliability for two judges is to simply correlate the judges scores across individuals using a Pearson=s correlation coefficient.

$$r = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{N \sum X^2 - (\sum X)^2} \sqrt{N \sum Y^2 - (\sum Y)^2}}$$

Here  $X_1$  through  $X_N$  are scores given to individuals 1 through N by judge X, and  $Y_1$  through  $Y_N$  are scores given to individuals 1 through N by judge Y. We do not expect users of the **PJRF** to compute correlations under all circumstances. If one is performing research and needs to show that Inter-rater reliability has been established, then it would be wise to report a Pearson=s correlation coefficient. Most statistical software packages for Windows will easily perform this calculation directly from a pull down menu, or the above formula can be used for calculations by hand. A statistical software package will quickly compute all pair-wise correlations between multiple raters.

The Pearson=s correlation coefficient tells us if there is a linear relationship between the scores given by the two raters. However, a high correlation does not mean

that one rater=s scores can be substituted for another rater=s scores. *A high linear similarity.* This is to say that there is agreement about who should get the highest score, who should get the second highest score,..., and who should get the lowest score. The Pearson=s correlation coefficient does not provide information about the relative stringency of raters. It is possible that one rater may be consistently more stringent in his or her judgments than another. Another characteristic of Pearson=s correlation coefficient is that it does not provide information about the spread (variance) of the ratings. These attributes should become clear with the following example.

The following table shows the scores given by rater X and rater Y for each of 5 individuals.

Individual	Score given by judge X	Score given by judge Y
1	15	9
2	14	7
3	13	5
4	17	13
5	16	11

As you can see the scores are different, yet the Pearson=s correlation coefficient

*correlation means that the raters are ranking individual*

between the two raters= scores is a perfect 1.00.

To reiterate, the reason that the Pearson=s product moment correlation between these two judges is 1.00 is that there is a perfect linear relationship between the scores of the two judges. Knowing that a strong linear relationship exists between the two judges is a very useful first step in establishing inter-rater reliability, but additional information is needed.

Taking a closer look at the scores of the two judges, one will notice that the average score given by judge X is 15 and the average score given by judge Y is 9. One will also notice that the scores given by judge Y are more spread out than the scores given by judge X. Using the suggested scoring scale provided on page 20, we see that judge X on average gives a marginal score and her scores range from positive to very strong. We also see that and judge Y on average gives a positive score and her scores range from negative to positive. It is of great concern that judge X could rate an individual positively while judge Y rates the same individual negatively (as is the case with individuals 2 and 3).

Along with the Pearson=s correlation coefficient, researchers should show that average scores given by judges are not significantly different and that the spread of scores given by judges are not significantly different.

In the above example Judges X and Y need to work on their understanding of the construct together and establish better agreement between themselves. We describe this process in detail in the following sections of this manual. If even after much work one judge still has a tendency to rate individuals higher or lower than another judge, additional steps can be taken. We suggest contacting the California Academic Press or an assessment expert or statistician on your site for further help on calibrating your judges. One may also wish to know how well two judges agree in their ratings of individual items on the **PJRF**. One can calculate the percentage of agreement between two raters on an individual=s form by using the following formula:

$$\frac{\text{Total \# of items where both judges agree}}{\text{Total \# of items on the scale}} * 100$$

To use this formula, divide the total number of agreed upon items (both raters answered yes or both raters answered no) by the total number of items (20 in the case of the **PJRF**) and then multiply by 100. This statistic is known as *point to point* agreement. Using point to point agreement in this manner only provides information about how raters agreed on a single individual, so it should only be used in conjunction with a Pearson=s correlation coefficient.

It is important to note that point to point agreement can be systematically affected by chance agreement; if two judges were asked to fill out the **PJRF** randomly, it is likely that some of their ratings would still correspond with each other. When all judges give high ratings it is important to see that the judges rated individuals negatively on the same items; when all judges give low ratings it is important to see that the judges rated individuals positively on the same items.

The kappa statistic is a reliability estimate which corrects for this chance effect.

$$Kappa = \frac{\frac{A}{N} - \left(\frac{X_{pos}}{N}\right)\left(\frac{Y_{pos}}{N}\right) + \left(\frac{X_{neg}}{N}\right)\left(\frac{Y_{neg}}{N}\right)}{1 - \left(\frac{X_{pos}}{N}\right)\left(\frac{Y_{pos}}{N}\right) + \left(\frac{X_{neg}}{N}\right)\left(\frac{Y_{neg}}{N}\right)}$$

$X_{pos}$  = number of times rater X gave the individual a point according to the scoring of the PJRF

$Y_{pos}$  = number of times rater Y gave the individual a point according to the scoring of the PJRF

$X_{neg}$  = number of times rater X failed to give individual a point according to the scoring of the PJRF

$Y_{neg}$  = number of times rater Y failed to give the individual a point according to the scoring of the PJRF

$N$  = 20 = number of items on the PJRF

$A$  = the number of times rater X and rater Y agree on the scoring of an item on the PJRF

Like point to point agreement, the kappa statistic used in this manner only provides information about agreement on a single individual. Another shortcoming of the kappa statistic is that it will break down at the

extremes. This is to say that if judges score individuals very high or very low, the relative importance of agreement on negatively or positively scored items becomes much higher; if two judges each give an individual a score of 19 out of 20 on the **PJRF**, yet the judges mark the individual down on different items the kappa statistic will be equal to 0.00, but if they agreed about the item to be negatively scored the kappa would be equal to 1.00. We suggest using a combination of the Pearson=s correlation coefficient, point to point agreement, and kappa statistics.

We have attempted to be complete in our explanation of the above reliability estimates. The above explanations of problems that can arise from misuse reliability coefficients apply to other rating forms as well as the **PJRF**. Inter-rater reliability is established for a given set of judges on a specific tool. When comparing individuals rated by separate judges it is important to have a previously established level of inter-rater reliability. Kappa statistics, point to point agreements, Pearson=s correlation coefficients, and other estimates of inter-rater reliability provide a gage of how much weight should be given to the comparison; the larger the reliability estimate, the more weight we give to the comparison.

## **The Professional Judgment Construct**

A mathematical correlation or kappa agreement coefficient is not enough; two or more raters may have similarly poor ideas

about the professional judgment construct. The use of rating forms such as the **PRJF** depends on the training and judgment of the raters. Raters need to be keenly focused on the professional judgment construct, performance standards, and well-practiced in how to make consistently reliable judgments in order for even the most valid of all possible rating forms to work.

Rating forms are tools used to measure performance through the observation of behavior. They are a means by which a researcher can assign a numerical value to observed behaviors, thus allowing quantification and statistical analysis. Rating forms make explicit the standards to be used for performance evaluation, yet they leave room for judgment in their application. Unlike an odometer rolling off the miles a car travels, a rating form requires a human being to make a decision about the quality of what is being observed. Like an odometer which is calibrated to measure a standard unit, the mile, it is crucial that raters be familiar with the performance standards in order to avoid errors. One must pay attention to exactly which phenomena one is trying to evaluate.

The phenomenon in the case of the **PJRF** is critical thinking, as described in the theory section of this manual. It addresses both critical thinking skills and habits of mind. It does not address such things as job knowledge, loyalty, respect, and concern for the common good. This is not to say that these things are unimportant. Rather, they

are simply not the focus of this tool. To assess job knowledge, loyalty, respect, concern for the common good, etc. all on a single rating form seems over-ambitious, and would arguably compromise the quality of such a broad tool.

Keep in mind that critical thinking is not necessarily about getting the right answer, but about how one approaches the problem. Critical thinking is not about agreeing with the person in charge, nor is it about obtaining someone's specifically desired result. Critical thinking is about the process of decision-making and problem solving. A person's ethics or politics should not affect a judge's rating, nor should their race or gender, nor their job description or position in the organization. Whether or not one likes the person being evaluated should not effect one's evaluation. Among errors introduced by the judge, the most serious is leniency or generosity. Lee Cronbach, 1990<sup>21</sup>.

Cronbach provides some reasons for leniency. The rater may feel that by rating an individual poorly he or she may be admitting poor leadership. The judge may have a general liking for her associates. The rater may be uncomfortable justifying criticism. Cronbach states that whether rating someone positively or negatively, in general the observer's overall evaluation strongly influences rating on specific traits. The tendency to give uniform ratings across the board is called a halo effect.

We have no reason to suppose that a person's knowledge of facts, loyalty to the firm, respect for authority, popularity, good looks, or charming manners are indications of how well the person can think nor how disposed the person is to approach problems by thinking.

When it comes to rating professional judgment, those with many years of professional experience are familiar with examples of people whose judgment in the practice of their profession is excellent and others whose judgment is very poor. We know people who are skilled at thinking and highly disposed to approach problems fairly-mindedly and thoughtfully. At the same time, we know people whose response to problem situations is exactly the opposite.

We see people who let their emotions or biases run rampant, who do not attend to relevant evidence, seem unable to follow the logic of the situation or to evaluate the consequences of possible choices.

There are many ways to achieve acceptable levels of inter-rater reliability. The goal, remember, is to be reasonably confident that those who are doing the rating are, on balance, evaluating things the same way. Establishing a high degree of inter-rater reliability is accomplished by getting two or more raters together and getting them to discuss the construct in an attempt to reach agreement with each other. The simplest

way of establishing a high degree of inter-rater reliability is for two experienced professionals, using the **PJRF**, separately to rate three or four novice professionals for practice. The two raters should then compare how they rated each novice on each item. Whether or not the ratings on a given item agree, the two raters should discuss how they came to their judgments. What did they notice about the novice's work performance, behaviors, approaches to problem-solving, comments, etc. that lead each of them to their separate, but similar, judgments about that item? If the raters are in accord about what leads each of them to their separate positive or negative evaluations, this is a good indication that progress on achieving mutually reinforcing standards of judgment for future consistency is being achieved.

When the two raters disagree, a similar conversation is needed, but in this case the focus should be on whether each rater was fully aware of what the other rater had observed as well as about the extent to which one rater or the other considered those things sufficient to warrant the negative or positive rating given. If the two raters can reach an accord about how the novice should be rated on that item, this is a good indication that inter-rater reliability is beginning to be established.

The two raters should continue through the entire set of practice ratings not for the purpose of using them to judge the

novices, but for the purpose of reinforcing at every possible point the extent to which the raters can be consistent in what they see and how they evaluate its significance.

The same process can be expanded to include more than two raters. In doing so, fuller discussions about the points of agreement and disagreement are held, making the discussions even more reinforcing. As an accord is reached, inter-rater reliability becomes stronger. Consequently, during the actual evaluation of novices, raters can be confident that what one regards as positive or negative will be evaluated in a consistent manner by other raters who have participated in the process.

The **PJRF** allows raters to bring some consistency and order to these fundamental realizations about how people do and do not use reasoning effectively as they make judgments in professional settings. The key to doing this well is to establish a consistent standard. Raters who wish to use the **PJRF** properly will want to assure themselves and those whom they are evaluating that their ratings are not idiosyncratic, uninformed, biased, or otherwise out of synch with what other equally well-informed and experienced raters would be saying.

Ideally, multiple judges will rate each individual in order to assure that the evaluations are unbiased, consistent with one another, and in accord with accepted

standards as applied in past situations. For high-stakes assessment purposes, we recommend using multiple raters for each individual judged. Many times separate judges rate different groups of individuals out of necessity. It may be that one rater does not know all of the novice professionals well enough to rate each of them; it may be that the structure of the organization or conflicting work schedules make it impossible to have the same judges rate each individual. If different judges are rating separate individuals, then it is very important that inter-rater reliability has been established.

When multiple judges collectively rate each individual every effort should be made so that the judges will agree on the rating given to that individual. If at the end of discussion there is still disagreement between raters, then the judges scores should be averaged.

Raters should individually work toward becoming better judges. Establishing personal consistency in rating is a similar process to the one outlined above. Like a conscientious teacher carefully grading essay examinations, one comes to be able to make qualitative judgments rather reliably if one keeps good notes and if one takes the time to reconsider and compare cases. While this strategy is not theoretically as good as having two different people rate the same person=s performance using the **PJRF**, it can be reasonably reliable. A single rater working alone must have a valid

understanding of the judgment construct. He or she should be a person who has previously worked with other raters using the **PJRF** and participated in the conversations and practice ratings that initially established inter-rater reliability for a group of raters.

To work in isolation of others, never having discussed your ratings with others to get a measure of accord in your application of the tool, is not recommended for high-stakes assessment purposes. On the other hand, using this solo approach as a way of gathering information on the basis of which to coach a novice and provide informal advice and guidance in the person=s development would not be unreasonable.

## Uses of The PJRF

There are a number of ways the **PJRF** can be used to improve an organization. Thinking organizations require individuals that can think for themselves and, perhaps of equal or greater importance, can think with others. The **PJRF** is useful in the development and evaluation of both. The

four uses described briefly here show the versatility of the **PJRF**. It can be used as a training tool, an evaluation instrument, a tool to assess the effect of staff development programs, and a Human Resources research tool.

**Training:** Senior professionals can use the **PJRF** and the Theory section of this Rater=s Manual to guide novices in understanding what is expected of them as thinking professionals, whether working alone on a problem or project or with others. Rich conversations can arise simply out of asking

Notice that this first example of using the **PJRF** as a training tool for individual and group development did not involve actually evaluating anyone. In fact, the power of this first use derives from its potential to shape behavior. As the standards by which people are going to be judged become known, people tend to shape their behavior toward those standards. If people are going to be given higher marks for thinking than for not thinking, then that is even a stronger reason to let them know that this is going to happen. For if they are thinking persons, they will be all the more likely to want to be conscious of what=s involved in good critical thinking and seek ways to demonstrate that they have the valued skills and habits of mind.

**Individual and Group Evaluation:** The **PJRF** can be used, obviously, to evaluate individuals and to evaluate groups with regard to their critical thinking in professional

how each of the critical thinking habits of mind and each of the skills and sub-skills are manifested when groups or individuals set about to solve problems in their particular professional field and workplace. This mentoring use of the **PJRF** and this Rater=s Manual can be extended by using the items on the **PJRF** itself as discussion starters. For example, ACan we describe a set of circumstances within which being intolerant of potential solutions that happen to be inconsistent with company protocols would be a mark of good thinking?@

judgment contexts. This is a main use of this tool. The resulting scores can be used in a number of ways. They can be kept on file and used in conjunction with other indicators and evaluations as part of an individual=s or team=s formal performance ratings. They can be discussed with the individual or with the group and be used either formally or informally as data contributing, along with other data, to a full picture of how well a person or group is doing and where they might look for ways to improve. Some of those improvements might be attitudinal, as for example, in relation to the critical thinking habits of mind. Some might be in terms of the use of critical thinking skills, particularly the skill of self-regulation, which involves monitoring one=s work for possible errors and engaging in self-correction of one=s own reasoning, if errors are discovered. Going back and thinking things through a second time is often a good idea for novices, and for

groups that may have lost sight of their initial assumptions as circumstances change and problems become redefined.

**Program Assessment:** The third use of the **PJRF** is to assess the effectiveness of training programs. The outcome being sought, of course, is that the persons who exit the program should have strong critical thinking habits of mind, if not skills as well, in their approach to professional judgment situations. The actual skills are best measured using an objective tool like the **California Critical Thinking Skills Test**.<sup>18</sup> The **PJRF** has the value of linking the skills and the habits of mind in a rating form that applies to the specific performance setting for the new professionals. Thus this tool, used in program assessment, involves having those persons exiting the training see many, if any, come through with ratings that showed them weak, negative, or even ambivalent about using thinking to approach and solve problems in the practice of that profession.

**Human Resources Research:** The fourth use of tools such as the **PJRF** is to investigate questions about the relationship of critical thinking in professional judgment and job performance. For example, **PJRF** ratings on new employees can be compared to the success these persons experience on the job and the kinds of performance ratings their supervisors give them. In time one can learn the extent to which scores on the **PJRF** do or do not correlate with individual success

program, whether it be company or university based, be evaluated by trained, experienced, objective professionals or preferably persons other than their teachers, supervisors, or mentors. The scores of the individuals are less important in this application than is the mean (mathematical average) score of the entire group who are exiting the training program. One is looking for evidence that the trainees, as a group, have internalized the value of critical thinking in the given professional judgment context. There may be some who are better at it than others, but it would be a strange profession indeed that was content to

in the organization and with the overall success of a given team, group, or office. These findings can, in turn, influence an organization's employee recruitment, training, and performance evaluation processes.

## References

1. Facione PA, Facione NC, Giancarlo CA. The motivation to think in working and learning. *Defining Expectations for Student Learning*. E. Jones (ed.). San Francisco, CA: Jossey-Bass Inc. Forthcoming 1996.
2. American Philosophical Association. *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction*. 1990. ERIC document ED 315-423.
3. Jones EA, Hoffman S, Moore LM, Ratcliff G, Tibbetts S, Click BL. *National Assessment of College Student Learning: Identifying the College Graduates= Essential Skills in Writing, Speech and Listening, and Critical Thinking*. Washington, DC: National Center for Educational Statistics, US Department of Education, Office of Educational Research and Improvement; 1995. OERI publication NCES 93-001.
4. Facione NC, Facione PA. Externalizing the critical thinking in knowledge development and clinical judgment. *Nursing Outlook*. 1996, vol 44, 129-36.
5. Facione NC, Facione PA. Assessment design issues for evaluating critical thinking in nursing. *Holistic Nursing Practice*. 1996, vol 10, 44-51.
6. Facione PA, Facione NC, Giancarlo CA. Critical thinking disposition as a measure of competent clinical judgment: the development of the CCTDI. *Journal of Nursing Education*. 1994, vol 33, 345-350.
7. Facione PA, Giancarlo CA, Facione NC, Gainen J. The disposition toward critical thinking. *Journal of General Education*. 1995, vol 44, 1-25.
8. Benner P. *From Novice to Expert*. Menlo Park, CA: Addison Wesley Publishing. 1984.
9. Gambrell E. *Critical Thinking in Clinical Practice*, San Francisco: Jossey Bass. 1990.
10. Newell, A. *Toward Unified Theories of Cognition*, Cambridge, MA: Harvard University Press. 1990.
11. King PM, Kitchener KS. *Developing Reflective Judgment*. San Francisco, CA: Jossey-Bass, 1994.

12. Giancarlo CA, The ideal critical thinker: Development of an expert Q-sort prototype. Presented at the American Psychological Association Meeting, Toronto, Canada, 1996.
13. Facione PA, Giancarlo CA, Facione NC, "Are college students disposed to think?" Paper presented at the Sixth International Conference on Thinking, Boston, MA, July 1994, and the annual meeting of the American Educational Research Association, New Orleans, LA, April 1994. ERIC Doc: ED 368-311.
14. Facione NC, Facione PA. *Critical Thinking Assessment in Nursing Education: An Aggregate Data Analysis*. Millbrae, CA: The California Academic Press, 1997
15. Facione PA. *The California Critical Thinking Skills Test: Technical Report #2 -- Factors Predictive of CT Skills*. 1990. ERIC Doc: ED 327 550
16. Giancarlo CA. *Critical Thinking, Culture and Personality: Predicting Latinos= Academic Success*. Unpublished doctoral dissertation. Department of Psychology, University of California at Riverside, 1996.
17. Mancinelli I. *The Role of Critical Thinking and Economic Advances of Mexican Americans in Corporate America*. Unpublished doctoral dissertation. Department of Psychology, Saybrook Institute, 1996.
18. Facione, PA. *The California Critical Thinking Skills Test*. Millbrae, CA: The California Academic Press, 1990.
19. Facione PA, Facione NC. *The California Critical Thinking Disposition Inventory*. Millbrae, CA: The California Academic Press, 1992.
20. Salvia J, Ysseldyke J, *Assessment*, Boston MA, Houghton Mifflin Company, 1991
21. Cronbach, LJ *Essentials of Psychological Testing*, New York, NY Harper and Row Publishers, 1990.