

**Programmevaluationen als Stärkung der Wissensbasis des FTI-
politischen Governancesystems?
Befunde der Untersuchung von zwölf Jahren Evaluationspraxis in
Österreich.**

Günther Landsteiner

16. Jahrestagung der DeGEval, 23.9.2016

Übersicht

„Jede Evaluation hat uns etwas gebracht, mit einer Ausnahme.“ (Agentur)

*„Wir wissen viel zu wenig. Wir wissen Einiges, aber wir wissen zu wenig
systematisch.“ (Ministerium)*

*„...It has long been understood that how evidence factors into policy development is a
function of the multidimensional and non-rational dynamics of the policy
process.“ (Cousins 2006)*

- 1. Studie: Metaevaluation in Kombination mit
Nutzungsforschung**
- 2. Zentrale Ergebnisse: Nutzen inwiefern, wodurch, woraus?**
- 3. Wie weiter?**

Die Studie: Auftrag und Motivation

Studie: Nützlichkeit und Nutzen der Programmevaluationen im Bereich der österreichischen FTI-Politik. Metaevaluation der Programmevaluationen 2003-2014 (Landsteiner 2015).

Auftraggeber: Rat für Forschung und Technologieentwicklung (RFTE), Wien

Motivation: Vielzahl von externen Evaluationen einzelner Programme in einem kleinteiligen Governancesystem (aktuell 136 Programme), Zweifel an deren Wirksamkeit bzw. Sinnhaftigkeit

Zweck: Evidenzproduktion zur Nützlichkeit der externen Programmevaluationen im österreichischen FTI-Bereich (verankert im Arbeitsprogramm des RFTE)

Hauptadressat: RFTE, um auf Basis der Studie zu eigenen Schlüssen und Empfehlungen gelangen zu können;

Weitere AdressatInnen: alle Akteure des Feldes, die Erkenntnisse in Zukunft bei der Beauftragung und Durchführung von Evaluationen aufgreifen können (AuftraggeberInnen, EvaluatorsInnen)

Durchführung: Juli 2014 – Dezember 2015 (6 Mannmonate)

Die Studie: Anforderungen

- **Evidenzbasierte Einschätzungen** der Nützlichkeit der Programmevaluationen, Identifikation von etwaigem Verbesserungspotenzial
- Bezugnahme auf ein **allgemein anerkanntes internationales Wissenskorporus** über nützliche und qualitätsvolle Evaluation und Aspekte der Evaluationsnutzung
- **Summativer Zugriff auf die langjährige Evaluationspraxis** auf Bundesebene
6 Institutionen in 3 Teilbereichen der FTI-Governance, holistisch betrachtet
12 Jahre (1.1.2003 - 30.9.2014): Reflexion der Evaluationspraxis seit den Strukturreformen am Beginn der 2000er-Jahre, incl. Nachvollzug von Weiterentwicklungen und kumulativen Evaluationswirkungen
- Erarbeitung von konkreten **Empfehlungen**
- **Unabhängiger Evaluator** mit guten Kenntnissen von System und FTI-Evaluation
- **Gangbarkeit und Akzeptanz**
 - Anonymität
 - Einbeziehung von EvaluatorInnen und AuftraggeberInnen (potenzielle NutzerInnen)

Konzeptuelle Grundlagen

Metaevaluation

Systematische Analyse einer oder mehrerer Evaluationen, die der Bestimmung von Wert und Güte dieser Evaluationen gilt.

Grundlage: DeGEval-Standards, mit dem Hintergrund der JCEE-Standards (2006 [1994])

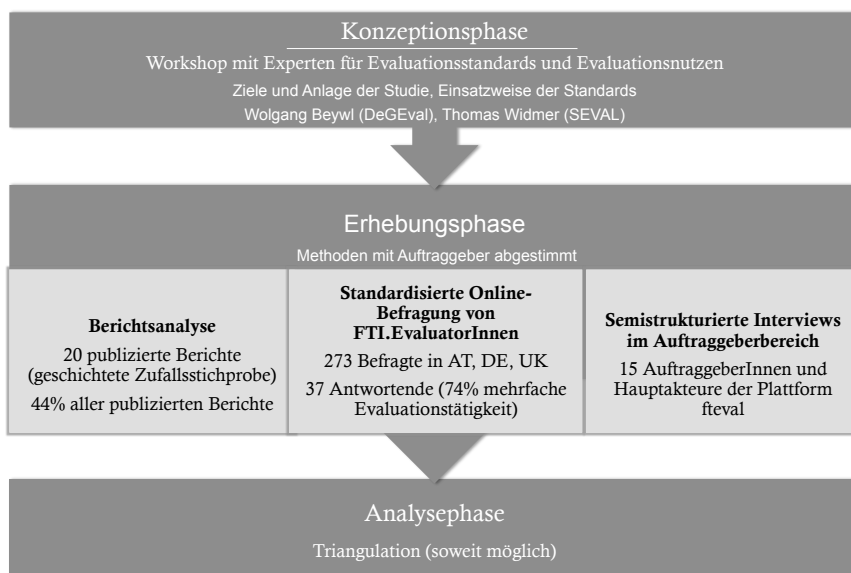
Gestaltungsweisen der Evaluationsberichte und der dahinter stehenden Evaluationsprozesse, die als essentielle Voraussetzungen für bzw. Ingredienzien der Nützlichkeit von Evaluationen zu erachten sind

Nicht: Auftragsbefreiung

Von Nützlichkeit zu Nutzen

- Nützlichkeit = Verkörperung von **Nutzungspotenzial**
 - Schaffung der Grundlagen für faktische Nutzungsweisen, analysierbar hinsichtlich von Stärken und Schwächen
 - Grundhypothese: Je nützlicher die Evaluationen in ihrer Planung, Durchführung und Präsentation sind, desto wahrscheinlicher wird eine tatsächliche Entstehung von Nutzen
- Entstehung von **faktischem Nutzen**: empirische **Nutzungsforschung**
 - tatsächlicher Nutzen kann durch weitere, externe Einflussfaktoren mit beeinflusst werden

Vorgehensweise



Vorgehensweise: Breite Analyse der Evaluationspraxis

11 DeGEval-Standards

- Alle Nützlichkeitsstandards soweit Information aus Berichten erwartbar, plus 5 (zunächst 4) Standards i.S. holistischer Evaluationsqualität: N1, N2, N4, N5, N6, N8, F3, F5, G2, G3, G8
- Berichte (Sample): Qualitative Sekundäranalyse (Rekursivität, Offenheit), Grading der Standard-Einhaltung zu Übersichtszwecken (5er-Skala), Dokumentation in factsheets incl. allgemeiner Charakterisierung der Evaluationen (Typ, Rolle form.-summ., Schwerpunkte n. OECD DAC)
- Standardisierte EvaluatordInnen-Befragung: Evaluationsprozesse, incl. 2 restliche N-Standards
- Semistrukturierte AuftraggeberInnen-Interviews: Evaluationsprozesse, incl. 2 restliche N-Standards

Typologie des Evaluationsnutzens (Patton 1997, Weiss 1998, Alkin/Taut 2003)

- Nicht: „evaluation influence“ (Kirkhart 2000, Henry/Mark 2003, Mark/Henry 2004)
- Direkte Operationalisierung in EvaluatordInnen-Befragung, interpretative Einordnung bei AuftraggeberInnen-Interviews

42 Einflussfaktoren auf die Entstehung von Evaluationsnutzen (Cousins/Leithwood 1986, Fleischer/Christie 2009, Johnson et al. 2009)

- Direkte Operationalisierung in EvaluatordInnen-Befragung, interpretative Einordnung bei AuftraggeberInnen-Interviews

Rolle evaluativer Information im Governancesystem, Einbettung der Evaluationsfunktion (Leeuw/Rist/Sonnichsen 2000, Rist/Stame 2006): AuftraggeberInnen-Interviews

Ergebnisse: entstandener Nutzen

- Die FTI-Programmevaluationen haben **Nutzen generiert**
- **Instrumenteller** und **konzeptueller** Nutzen dominieren
 - Entscheidungsfindungen unterschiedlicher Dimension auf unterschiedlichen Ebenen (Programmumsetzende – Programmeigentümer - Politik)
 - Lerneffekte auf unterschiedlichen Ebenen - teils von außen schlecht erkennbar, teils nicht an einer bestimmten Evaluation festzumachen
 - auch Evaluationen von nicht weitergeführten Programmen haben Nutzen erzeugt
 - hohe Variabilität (unterschiedliche Konfigurationen von Programmzuständigkeiten, Akteursbeziehungen, Aufmerksamkeiten)
- **Organisatorische Anpassungen** zum Umgang mit Evaluation und deren Ergebnissen wurden vorgenommen (unterschiedlich an verschiedenen Systemstellen)
- Erzielung von **symbolischem Nutzen** („Überzeugen, Rechtfertigen“) schwingt stark mit: Aushandlungsprozesse in Multi-Akteurs-Szenarien, Verkopplung von Lernorientierung mit Rechenschaftslegung (Schnittstellen-Funktion der Programmevaluationen im Governancesystem)
- **Prozessnutzen** und **Aufklärung** vorhanden, aber gering ausgeprägt
- Routine ohne Effekte existiert ebenfalls, aber von AuftraggeberInnen und EvaluatordInnen als nachrangig eingeschätzt
- (Mißbräuchliche Nutzung: nicht untersucht, anderes Studiendesign wäre gefordert)

Ergebnisse: entstandener Nutzen

- ◆ **Gemengelage** unterschiedlicher Nutzungstypen, verschiedene Nutzenarten schließen einander nicht aus
- ◆ **multiple Effekte** einzelner Evaluationen, kumulative Lerneffekte
- ◆ Programmevaluationen sind **konkurrenzloses Instrument** der Wissensproduktion im Governancesystem
- ◆ Programmevaluationen sind **nicht alleinige Grundlagen FTI-politischen Handelns**, explizite Vorbehalte gegen unmittelbare Nutzung (kein „sklavisches“ Befolgen von Empfehlungen)
- ◆ **Nutzungsstufen** Datenlage # Schlussfolgerungen und Empfehlungen
- ◆ Gemeinsamkeiten und Unterschiede der **politisch-administrativen Teilbereiche**
 - Programmtypen
 - Autonomiestatus der Agenturen, Entscheidungskompetenzen zu Programmen
 - Organisationskulturen, Personen, Netzwerke
 - Variierende politischen Aufmerksamkeiten und „Eigensinnigkeiten“

Ergebnisse: Einflussfaktoren auf Nutzenentstehung

Evaluationsmerkmale und Kontextfaktoren gleich stark wirksam

Top20 der EvaluatordInnen-Befragung, durch AuftraggeberInnen-Interviews weitestgehend bestätigt

- **Wesentlichste Evaluationsmerkmale** (gereiht gemäß EvaluatordInnen-Befragung):
 - Glaubwürdigkeit der EvaluatordInnen
 - Klarheit der Berichterstattung (auch: Executive Summary)
 - Zuschnitt auf Informationsbedürfnisse der AuftraggeberInnen
 - Stil der Berichterstattung
- **Wesentlichste Kontextfaktoren** (gereiht gemäß EvaluatordInnen-Befragung):
 - Erwartung der AuftraggeberInnen, dass ihnen die Evaluation von Nutzen sein wird
 - Konnex mit Entscheidungsdruck, aktueller Problemlage, „window of opportunity“
 - Personenabhängigkeit (persönliche Sichtweisen, Konsistenz damit, Erfahrung, Rolle in Organisation)
 - Organisationsmerkmale (Ressourcen, Erfahrung, organisatorische Anpassungen)
- **Methoden?**
 - Stellenwert von Methodenmix, Triangulation, Finesse der Methodenanwendung recht untergeordnet (EvaluatordInnen)
 - Methodische Mängel haben die Nutzbarkeit einzelner Evaluationen eingeschränkt, ihre Publikation verhindert (AuftraggeberInnen)
 - „triviale Ergebnisse“, „Charakter des Programms nicht gut getroffen“(AuftraggeberInnen)

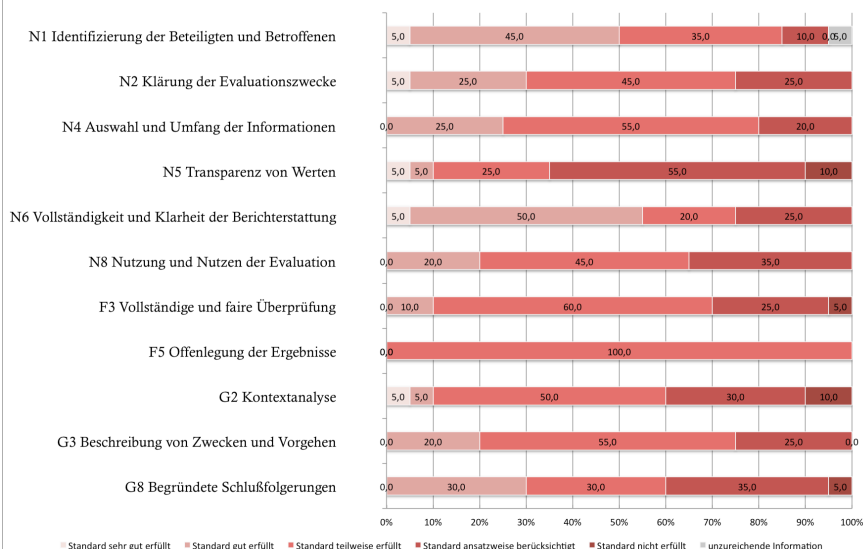
Ergebnisse: Berichtsanalyse übergreifend

- Zwischenevaluationen mit Multi Purpose-Charakter dominieren klar
95% Zwischenevaluationen oder Teile von Begleitevaluationen, 70% formativ-summativ, 60% drei oder mehr Evaluationsschwerpunkte
- Programmevaluationen in Anlage, Durchführung und Berichterstattung recht unterschiedlich (individuelle Profile - Erfüllungsniveaus von Standards streuen)
- Einschätzungen zu einem Standard implizieren oft mehrere Gesichtspunkte, Stärken können durch Schwächen konterkariert sein
- Mittleres, grundsätzlich zufriedenstellendes Niveau (5stufige Skala) dominiert
 - beste und schlechteste Einstufung kaum vergeben; bereits in früheren Phasen gute Evaluationen/Berichte, auch in jüngster Zeit merkliche Schwächen
 - Steigerungstendenzen über die Zeit bei der Erfüllung der meisten der 11 Standards, aber auch persistente Herausforderungen

Ergebnisse: Einstufung der analysierten Berichte

Einstufung der analysierten Berichte nach herangezogenen Standards (% der Berichte)

Erfüllungsniveaus der herangezogenen Standards (% der 20 analysierten Berichte)



Ergebnisse: Symptomatik anhand der herangezogenen Standards

Daten: Berichtsanalyse, EvaluatorsInnen-Befragung, Auftraggeber-Interviews

N1 Gezielte Einbeziehung von Lernpartnern bzw. frühe Abklärung von Informationsbedarf mit Stakeholdern gering, kaum Einbeziehung von Zielgruppen in die Evaluationsplanung oder von nicht unmittelbar evaluationszuständigen politisch-administrativen Akteuren, die am evaluierten Programm bzw. den Evaluationsergebnissen Interesse haben könnten.

Kontext: Evaluationsverantwortlichkeit an Zuständigkeit für die Programme gebunden, kleinteiliges bzw. zersplittertes System. Siloisierung der Evaluationsplanung. Abstimmungen beruhen stark auf Engagement und Kooperationsbereitschaft.

N2 75% aller analysierten Berichte geben Lernorientierung zu erkennen. Evaluationszwecke allerdings i. d. R. multipel, nur kursorisch umrissen bzw. unscharf.

Nicht immer klare Arbeitsaufträge für EvaluatorsInnen (39% „immer“). Nicht immer Priorisierung von Hauptzweck(en) (29% EvaluatorsInnen „immer“)

Kontext: Schnittstellenfunktion der Programmevaluationen zwischen Ressorts und BMF: Rechenschaftslegung und symbolischer Nutzen („Überzeugen“) schwingt immer mit. Vorgegebene Evaluationszeitpunkte und –budgets.

Evaluationsvorgaben (zeitlich, inhaltlich, finanziell) *versus* aktueller Erkenntnisbedarf.

N3 Glaubwürdigkeit der EvaluatorsInnen ihnen zufolge essentiell.

Kontext: Kontinuierliche Kommunikation in Plattform fteval, Beauftragung FTI-spezialisierter Institute; wiederholte Tätigkeiten/Beauftragungen von EvaluatorsInnen

Ergebnisse: Symptomatik anhand der herangezogenen Standards 2

N4 Versammlung von Outputs, Outcomes und ersten Impacts bei nicht klar ausgewiesenen Fragestellungen; Auseinandersetzung mit Programmanlagen und –umsetzungen ausschnitthaft; Unklarheiten mit Zielsystemen und Status von Beobachtungen als Outputs/Outcomes, intendierte Effekte/Nebenwirkungen; Unklarheiten zu direkten/indirekten Zielgruppen und deren Erreichung/Beeinflussung.

Tendenz „Black Box“ bezüglich Wirkungsweisen, teils Züge eines „evaluability assessment“.

AuftraggeberInnen: Erwartungen ex-post oft unrealistisch. EvaluatorsInnen: gezielte Schritte zur Abstimmung der Erwartungen 29% „immer“.

Evaluationsdesign erlaubt gute Behandlung aller Evaluationsfragestellungen: 23% EvaluatorsInnen „immer“.

Eingeplante Flexibilität, um während der Evaluationsdurchführung auf veränderte Umstände reagieren zu können: 18% EvaluatorsInnen „immer“.

Know-How der EvaluatorsInnen hat Einfluss auf Evaluationsfragestellungen: 26% EvaluatorsInnen „immer“.

Kontext: Vorgegebene Evaluationszeitpunkte, -budgets und –fragestellungen; kurze Vergabeverfahren mit TORs; Ressourcenbeschränkungen der AuftraggeberInnen für Auseinandersetzung mit der Evaluation; persönliches Engagement nach Maßgabe der Möglichkeiten

Ergebnisse: Symptomatik anhand der herangezogenen Standards 3

- N5 Bewertungsmaßstäbe oft wenig geklärt, Kriterien für Ergebnisinterpretation kaum vorab festgelegt. Einzelaspekte entlang „üblicher“ Betrachtungsweisen; Abwägungen zu Bewertungsfragen „mit offenem Ausgang“.
- Bewertung als reine EvaluatorInnen-Aufgabe, unterschiedliche Vorgehensweisen der EvaluatorInnen (im Sinne der Auftraggeber/ohne Rücksicht auf Auftraggeber-Sichtweisen).
- N6 Nicht immer stringent-logischer Berichtsaufbau, versteckte Schlussfolgerungen und Empfehlungen, Unklarheiten zu Programm und Analytik, Variabilität der Executive Summaries.
- N7 Bezug zu anstehender Entscheidungsfindung laut EvaluatorInnen wesentlicher Einflussfaktor für Nutzenentstehung.
- Kontext:** Aktuelle FTI-politische Themenstellungen oder „Windows of Opportunity“ kaum Auslöser von gezielten Evaluationsaktivitäten. Vorgegebene Evaluations- und Verhandlungsrhythmen für formale Programmvereinbarungen dominieren.
- N8 Kaum gezielte Planungen bezüglich Art und Entstehungsweise des Evaluationsnutzens.
- Kontext:** Bisweilen geringe Nutzenerwartungen der AuftraggeberInnen bzw. HauptadressatInnen, „Pflichtübung“.

Ergebnisse: Symptomatik anhand der herangezogenen Standards 4

- F3 Stärken- und Schwächen-Abschnitte der Evaluationsberichte, Unklarheiten welche Bedeutung Datenlagen zugemessen wird, Faktenbezug durch Hintergrundwissen und Annahmen der EvaluatorInnen ergänzt, manchmal unmittelbar wertende Sprache.
- F5 Publikation ist Usus, aber unbekannt Zahl von Ausnahmen. Vollständige Publikation des Originalberichts ohne Eingriffe nicht immer der Fall (insbes. Datenlagen in Anhängen). Evaluative Information nur bedingt gezielt an Stakeholder herangetragen - publizierte Information bleibt Holschuld für vorinformierte InteressentInnen.
- Auseinandersetzung mit den Evaluationsergebnissen beschränkt sich im Wesentlichen auf die wenigen unmittelbar administrativ Zuständigen, sonst unsystematisch bzw. akzidentuell.
- Kontext:** Keine systematischen Vorkehrungen für Kenntnisnahme der evaluativen Information im Governancesystem. Informelle Kommunikationsnetzwerke, persönliches Engagement, Ressourcenproblematik.
- G2 Kontextualisierung wird in unterschiedlicher Weise versucht, stets deutlich eingeschränkt.
- AuftraggeberInnen: klarer Bedarf an systemisch orientierten Erkenntnissen (Positionierung der Maßnahme, Steuerungsmöglichkeiten im Maßnahmenportfolio)

Ergebnisse: Symptomatik anhand der herangezogenen Standards 5

- G3 Einschätzbarkeit der Gesamtvorgehensweise der Evaluationen mit ihren Vor- und Nachteilen deutlich beeinträchtigt. Gesamtkonzepte der Analytik gegenüber Programm kaum greifbar. Datenbeschränkungen (Verfügbarkeit, Tragfähigkeit) nicht immer berichtet, Evaluationsprozesse (Einbeziehungen, Entscheidungen, Organisatorisches) selten systematisch dargestellt.
- G8 Schlussfolgerungen und Empfehlungen mit Spielraum gegenüber Gesamtgehalt des Berichts. Überschneidung mit ExpertInnengutachten: Datenlage und –interpretation *versus* persönliches Hintergrundwissen, Sichtweisen.

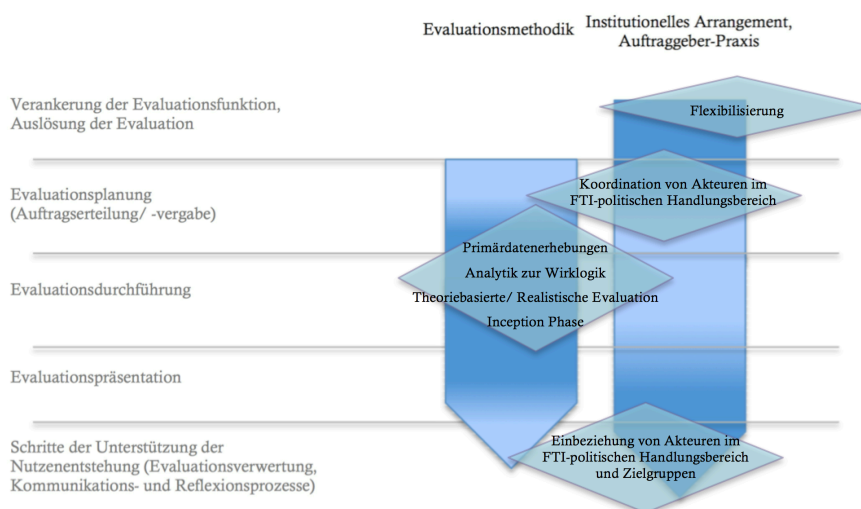
Wie weiter?

- ◆ Gemengelage von Nutzentypen, Bündel von evaluationsinhärenten und externen Faktoren von hoher Komplexität – kein Hauptfaktor als „Stellschraube“
- ◆ Konzeptueller Nutzen (Lernen) entsteht unmittelbar genauso wie mittel- bis längerfristig und kumulativ – aber „gefangen“ in kleinteiligen Silos (Principal-Agent-Verhältnisse, nicht immer tariert)
- ◆ Art der Verankerung der Evaluationsfunktion im Governancesystem schlägt durch auf Auslösung, Ressourcenausstattung und Planung der Programmevaluationen
- ◆ Organisationseigenschaften und ‚Human Factor‘ im Auftraggeberbereich bestimmen Möglichkeiten der Evaluation und ihres Nutzens entscheidend mit
- ◆ Evaluationsnutzen entsteht auch auf der Grundlage von nicht optimalen Grundlagen im Sinn der Standards
 - ➔ Entstehung von Evaluationsnutzen überhaupt *versus* Schaffung bestmöglicher Grundlagen für größtmöglichen Evaluationsnutzen (*evidence-based policy-making versus policy-based evidence-making*, Strassheim/Kettunen?)
- ◆ Verantwortlichkeiten für Evaluationsqualität und Nutzenentstehung verteilen sich auf EvaluatorInnen und AuftraggeberInnen gleichermaßen (Standards, Nutzungsforschung).

Wie weiter?

- ◆ Lernfunktion der Programmevaluationen stärken: Flexibilisierung
- ◆ Prozessqualität der Programmevaluationen sichern
- ◆ Comprehensive *versus* tailored evaluation (Lösung vom Schematismus)
- ◆ Organisation von Reflexion im verteilten Governancesystem
- ◆ Positionierung der EvaluatordInnenrolle?
 - Größtmöglicher Nutzen, entstandener Nutzen als zentrales Qualitätskriterium (Patton)
 - Objektivistische Evaluationsqualität unter starker Beachtung der Bewertungsproblematik (Scriven, „alpha science“)
 - Multidimensionale Evaluationsqualität unter Anerkennung der Grenzen für bzw. Volatilität von Nutzenentstehung (Weiss)
- ◆ Verbesserungsmöglichkeiten im gegebenen Arrangement (20 Empfehlungen)
 - Evaluationsmethodik
 - Institutionelles Arrangement und Praxis im Auftraggeberbereich
- ◆ einzelnes Evaluationsprojekt inwieweit abhängig von institutioneller Verankerung der Evaluationsfunktion?

Wie weiter? Einige Ansatzpunkte ...



Vielen Dank für Ihre Aufmerksamkeit!

Studie unter: www.rat-fte.at/publikationen.html#2015

MMag. Günther Landsteiner
Research-Consultancy-Evaluation
landsteiner-rce@chello.at