

Curso: MBA 2014

Hacia un Nuevo Enfoque en Evaluaciones de Desempeño para Equipos de Desarrollo de Software Ágil en Argentina.

Alumno: Ricardo A. Ferrari

Tutora: Liliana Moya

Lugar: Ciudad Autónoma de Buenos Aires

Fecha: 23 de abril de 2019

## Agradecimientos

Este trabajo ha demandado el apoyo y la colaboración de muchísimos compañeros muy valiosos. A riesgo de pasar por injusto debido a algún olvido involuntario, optaré por reconocer a varios con nombre y apellido: Gabriel Gavinowich, Gabriel Vinokur, Pablo Tortorella, Adrián Valerio Anacleto, Sofía Poncino, Marcelo Ferreiro, Javier Bruges, Dan Rozenfarb, Gerardo Zmener, Diego Gutiérrez, Luis Introcaso, Andrés Garagiola, Nicolás Dimov, Ezequiel Turovetzky, Mariano Goldsman, Ramiro Rela, Pablo Zivic, Gabriel Antrilli, Alfonso Bistakis, Juan Martín Bonacci y Andrea Korniusza. Todos ellos han aportado tiempo, predisposición, aliento e ideas, materiales escasos entre tanta urgencia cotidiana.

Agradezco al plantel docente de la Universidad Torcuato Di Tella, de quienes he aprendido muchísimas cosas nuevas. Resalto que Liliana Moya ha hecho grandes contribuciones al presente trabajo. Muchas de las virtudes que el lector encuentre en él hubieran sido imposibles sin su atención y profesionalismo. Las flaquezas y puntos bajos, al contrario, corresponden solamente al autor.

Entre mis buenos jefes, aprecio el apoyo de la dirección de Olenick & Associates, que me ha provisto de herramientas para realizar este trabajo. También quisiera destacar a Cristián Argüello, que logró transmitirme la importancia de seguir buenas prácticas y ha sido un valioso ejemplo en lo profesional y personal.

El presente trabajo es un mojón en una trayectoria académica que concibo como un proceso de construcción colectiva. Por eso, también es fruto del aporte de mis compañeros académicos de la Escuela Técnica ORT 1, la Facultad de Ciencias Sociales de la Universidad de Buenos Aires y

(desde ya) del MBA Vespertino 1 2014 en esta casa de estudios. Me han inspirado, desafiado y obligado a mejorar: también a ellos les debo gratitud.

Por último, agradezco el apoyo de mi familia, mis amigos y mi compañera María Laura Guevara. Sin su apuntalamiento y comprensión este trabajo hubiera continuado demorándose, quizás al punto de no escribirse jamás.

## Resumen

La adopción de Metodologías Ágiles en desarrollo de software ha conllevado una serie de cambios en la dinámica y estructura de los equipos. El presente trabajo busca comprender cómo se vincula esa nueva realidad con las Herramientas de Evaluación de Desempeño y qué tienen estas últimas para aportarle a las empresas que producen aplicaciones o tercerizan servicios de programación. Se caracterizará la teoría sobre dichas herramientas, como también la propuesta de las Metodologías Ágiles y qué métricas se pueden recabar a través del proceso. Posteriormente, se presentarán los hallazgos de estudios de caso respecto a la interacción entre ambos aspectos. Por último, se presentarán recomendaciones que busquen guiar la implementación de Herramientas de Evaluación de desempeño y la práctica cotidiana de desarrollo siguiendo Metodologías Ágiles para optimizar su rendimiento.

## Palabras Clave

Evaluación de Desempeño, Gestión del Desarrollo, Gestión del Talento, Metodologías Ágiles, SCRUM, Desarrollo de Software, Control de Calidad, Buenos Aires, Argentina.

## Índice

Introducción .....	9
Cuerpo Teórico .....	13
Capítulo I: Herramientas de Evaluación de Desempeño .....	14
I.A. Sus Objetivos .....	14
I.B. Métodos de Evaluación .....	20
I.C. Participantes del Proceso.....	30
I.D. Características, Desafíos, Recomendaciones y Limitaciones .....	34
Capítulo II: Desarrollo de Software Siguiendo Metodologías Ágiles .....	40
II.A. Dinámica General .....	41
II.B. Métricas.....	47
II.C. Métricas y Evaluaciones: cómo evitar que devoren al equipo.....	54
Cuerpo Empírico.....	60
Capítulo III: La relación práctica entre las Herramientas y los Equipos .....	62
III.A. Metodología y Participantes de la Muestra .....	62
III.B. Caracterización de los Respondientes (por Rol y Tamaño de Empresa).....	62
III.C. Herramientas de Evaluación del Desempeño .....	65
III.C.I. Frecuencia de Reuniones de Feedback y Métricas Capturadas .....	65
III.C.II. Métodos y Calibración .....	68

III.C.III. Aplicación de los Resultados .....	72
III.C.IV. Participación de la Dirección.....	75
III.D. Metodologías Ágiles.....	77
III.D.I. Metodologías Ágiles: Una Advertencia .....	78
III.D.II. Métricas Relacionadas con la Metodología y el Negocio.....	80
III.D.III. Métricas Relacionadas con Web/App Development .....	84
III.D.IV. Métricas Relacionadas con Calidad del Código .....	87
III.D.V. Reuniones Retrospectivas: ¿Cómo se consolida esa información? .....	90
III.E. Percepción General del Sistema .....	92
III.E.I. Grados de Satisfacción con la Herramienta Actual .....	92
III.E.II. Peso Práctico .....	95
Capítulo IV: Hacia una Implementación Satisfactoria .....	100
Conclusiones.....	109
Bibliografía .....	113
Apéndices / Anexos .....	118
Apéndice I: Cuestionario Guía para las Entrevistas.....	118

## Índice de Cuadros y Gráficos

Figura 1 – Caracterización de los Respondientes (por Rol) .....	63
Figura 2 - Existencia de una Herramienta de Evaluación del Desempeño en Funciones .....	64
Figura 3 - Frecuencia de las Reuniones de Feedback .....	65
Figura 4 - Métricas/Herramientas Relevadas Individualmente .....	67
Figura 5 - Metodologías Usadas en su Herramienta de Evaluación .....	69
Figura 6 - Herramientas para Garantizar Dispersión de Calificaciones.....	71
Figura 7 - Modos de Calibración .....	72
Figura 8 - Aplicaciones de la Información Individual Provista por la Evaluación de Desempeño .....	74
Figura 9 - Acciones en la que Participa la Dirección.....	76
Figura 10 - Métricas Relacionadas con la Metodología y el Negocio (Sólo Grupales) .....	81
Figura 11 - Métricas Relacionadas con la Metodología y el Negocio (Individuales/Grupales) ..	83
Figura 12 - Métricas Relacionadas con Web/App Development.....	84
Figura 13 - Métricas Relacionadas con la Calidad del Código.....	87
Figura 14 – Percepción de la Eficiencia con que la Herramienta de Evaluación mide Objetivos Individuales.....	93
Figura 15 - Percepción de la Eficiencia con que la Herramienta de Evaluación mide Potencial Individual .....	94
Figura 16 - Percepción de la Eficiencia con que la Herramienta de Evaluación mide Colaboración con el Equipo.....	95
Figura 17 - Percepción de Mayor Peso Práctico (Recategorizaciones Individuales) .....	96

Figura 18 - Percepción de Mayor Peso Práctico (Retribución Salarial Individual)..... 98

Figura 19 - Percepción de Mayor Peso Práctico (Planes de Carrera Individuales) ..... 99

## Introducción

Durante los últimos años, una serie de cambios organizacionales a nivel global han alentado transformaciones en los Sistemas de Gestión del Desempeño. La ya consolidada economía digital ha impulsado cambios en la estructura organizacional (ahora más tendientes a la horizontalidad), ha generado nuevos comportamientos por parte de los empleados y del mismo modo ha impulsado una valoración diferente de las habilidades de las personas. En consecuencia, la capacidad de los individuos de desarrollarse, colaborar con sus equipos y participar no solamente en tareas rutinarias sino también en planificación y creación de las políticas corporativas se ha convertido en un atributo especialmente valorado.

Las Herramientas de Medición Individual del Desempeño continúan siendo instrumentos vitales para alinear objetivos y comportamientos de los empleados con las necesidades del negocio. No es sorprendente que su uso siga siendo extendido a nivel global. No obstante, la información que proveen y su eficacia sigue estando puesta en entredicho: se las percibe como mucho más eficaces para satisfacer los objetivos propuestos por el Performance Management tradicional (por ejemplo, alcanzar una calificación general de los empleados para identificar el nivel de contribución de cada uno, definir una remuneración y tomar decisiones de carrera) que para contribuir al desarrollo de capacidades de los empleados. En este contexto, es lógico que los vientos de cambio que soplan sobre los Sistemas de Gestión de Desempeño preanuncien transformaciones en las herramientas de las cuales dependen.

La industria del software no es ajena a estas transformaciones, y quizás sea aquella en la que las modificaciones se evidenciaron con mayor velocidad. La súbita popularidad de las Metodologías Ágiles de Desarrollo de Software fue al mismo tiempo una respuesta desde una

perspectiva de procesos (porque permitió gestionar equipos donde la responsabilidad es colectiva, la estructura es horizontal y el *feedback* frecuente con el cliente es esencial para el éxito) como también desde una óptica de rentabilidad: al proponer entregas frecuentes e incrementales, hay elementos para medir costos y plazos y accionar rápidamente en caso de que se presenten desvíos de cualquier tipo.

El advenimiento de estas metodologías ha puesto en el candelero a todo tipo de habilidades de gestión e interacción entre el equipo de trabajo y sus clientes finales. Además, ha forzado a los desarrolladores de software a no ser meros implementadores de diseños ajenos. Para tener éxito a largo plazo, los equipos de desarrollo deben hacer diseños robustos, comprender las necesidades del cliente, hacer implementaciones mantenibles, ofrecer soluciones elegantes y cooperar de modo permanente. Dada la importancia de asegurar el avance y la cooperación entre los miembros del equipo, la metodología demanda una periodicidad de comunicación al menos diaria para confirmar prioridades y que se asegure que cada una de las tareas tiene responsables asignados.

En consecuencia, cualquier equipo de desarrollo de software que sigue Metodologías Ágiles genera (en beneficio propio) un inmenso volumen de información que es accesible a todos los miembros. El procesamiento diario de esos datos es requisito para medir progresos, introducir mejoras o activar planes de contingencia si fuera necesario, con lo cual el equipo es plenamente consciente de su posición (como grupo y como miembro individual) a cada momento.

Este trabajo busca responder al siguiente interrogante: ¿cómo se integran el volumen de información cotidiana que se genera cuando un equipo trabaja siguiendo metodologías ágiles y aquella que se captura a través de las Herramientas de Evaluación de Desempeño? Incluso, subyace la pregunta de si estas últimas continúan teniendo una razón de ser: eso sólo tendría sentido

práctico si tuvieran algo sustancioso para aportar sobre el rendimiento de cada uno de los miembros de un equipo de desarrollo de software.

Se busca abordar esta cuestión a través de un enfoque teórico y práctico. Se parte del desarrollo conceptual detrás de las Herramientas de Evaluación de Desempeño y el desarrollo de software siguiendo Metodologías Ágiles, y luego se estudia cómo estos aspectos se articulan en la práctica en 14 empresas con sede en el Área Metropolitana de Buenos Aires.

El marco teórico abarca los Capítulos I y II.

En el Capítulo I se caracteriza a los diferentes actores y procesos que constituyen a las Herramientas de Evaluación de Desempeño, tanto tradicionales como más modernas. Se presentan los desafíos más frecuentes para cada una de ellas y las limitaciones prácticas que suelen surgir.

En el Capítulo II se analiza a las Metodologías Ágiles de desarrollo de software, dando cuenta de su dinámica cotidiana, sus fundamentos, del modo en que contribuyen a construir productos alineados con los deseos del cliente, cómo ellas ayudan a mejorar los procesos seguidos por el equipo y las métricas (tanto individuales como colectivas) que permiten medir avances y grados de éxito.

El Capítulo III permite entender la vinculación práctica entre las Herramientas de Evaluación de Desempeño y los Equipos de desarrollo de software que siguen Metodologías Ágiles. En él se cristaliza un cuerpo empírico compuesto del análisis de entrevistas realizadas a miembros de 14 empresas de software (gerentes, representantes de Recursos Humanos, Líderes de Equipos, Desarrolladores). Si bien dichas entrevistas siguen un cuestionario guía, se ha recabado y analizado información diversa según la especificidad de cada una de las empresas relevadas.

El fruto de dichas entrevistas permite proponer respuestas a los siguientes interrogantes:

- ¿De qué modo evalúan las empresas argentinas el rendimiento individual de los desarrolladores de software que trabajan siguiendo Metodologías Ágiles?
- ¿Qué razones se reconocen para haber elegido una determinada herramienta de evaluación? ¿Se han tenido en cuenta las características de la empresa y la naturaleza de sus equipos de desarrollo?
- ¿Qué métricas forman parte o complementan dicha evaluación?
- ¿Cómo se realimenta la herramienta de evaluación de desempeño individual con la información que provee la propia dinámica de las Metodologías Ágiles (en especial a través de demostraciones al cliente y reuniones retrospectivas del equipo)?
- ¿De qué modo se utilizan los resultados de las evaluaciones de desempeño?

En el Capítulo IV se presenta una guía para lograr una relación más fluida entre las Herramientas de Evaluación de Desempeño y los equipos de desarrollo de software, incluyendo recomendaciones para sacar el máximo provecho de la información cualitativa y cuantitativa que los equipos producen permanentemente cuando siguen Metodologías Ágiles. Dicho conjunto de propuestas analiza la viabilidad de diferentes métricas para relevar y alinear los objetivos y comportamientos de los evaluados con los intereses de su equipo, y facilita la identificación de claves para el desarrollo individual.

Por último, las conclusiones sintetizan las observaciones hechas durante los estudios de casos prácticos del capítulo III y las recomendaciones desarrolladas durante el capítulo IV, buscando delinear las principales ventajas de seguir evaluaciones de desempeño basadas en objetivos y comportamientos y fortalecidas con el análisis de métricas del área de desarrollo, del área de calidad y aquellas de la propia Metodología Ágil.

## **Cuerpo Teórico**

Para conseguir describir y analizar cómo se interrelacionan el volumen de información capturado por las Herramientas de Evaluación de Desempeño (generalmente a nivel individual) y el que se desprende de las Metodologías Ágiles (tanto a nivel individual como grupal) es preciso presentar al lector los usos más frecuentes de cada una de ellas. Esto incluye las razones de su existencia, el modo en que contribuyen a que una empresa alcance sus propósitos estratégicos, los objetivos que persiguen, los actores cotidianos que cimientan su dinámica, los desafíos que cada una de ellas conlleva y el modo en que se proponen superarlos. El Cuerpo Teórico del presente trabajo hará foco esos aspectos en los siguientes Capítulos I y II, abordando en el primero de ellos a las Herramientas de Evaluación de Desempeño y en el segundo a las Metodologías Ágiles de Desarrollo de Software.

## Capítulo I: Herramientas de Evaluación de Desempeño

En el presente capítulo se busca caracterizar a las Herramientas de Evaluación de Desempeño. En primer lugar, se explicita qué propósito buscan servir a una empresa, qué explica su valor final y cuál es su relación con la gestión del talento. Luego, se mencionan los métodos más extendidos de evaluación, presentando virtudes y defectos de cada uno. Posteriormente, se caracteriza a los participantes del proceso y sus experiencias. Por último, se detallan las dificultades más comunes que amenazan los resultados del proceso y las recomendaciones que maximizan las chances de tener éxito.

### I.A. Sus Objetivos

Puesto que la productividad y rentabilidad de las empresas depende directamente de la performance de sus empleados, es lógico que una gran mayoría de ellas implemente Sistemas de Gestión del Desempeño y se apoye en Herramientas que buscan operacionalizar el rendimiento individual. Es intuitivo coincidir con el siguiente precepto: es imprescindible medir para gestionar. Así, continúa habiendo consenso en que las evaluaciones individuales de desempeño son una herramienta importante para que los ejecutivos, jefes, gerentes y supervisores consigan alinear la performance de los empleados con los objetivos de la empresa.

La primera gran advertencia es que para que dichas evaluaciones sean exitosas, deben primeramente estar bien definidas la estrategia corporativa e identificadas las ventajas competitivas de cada organización.<sup>1</sup> Si se aspira a tener éxito, no se tratará solamente de medir para gestionar, sino de medir sólo lo que es relevante para la organización.

---

<sup>1</sup> Para un mayor desarrollo sobre esta idea y casos de estudio al respecto, véase Hatum (2009), pp. 107-108 y sigs.

Una vez que el pilar estratégico está establecido se vuelve palpable que hay múltiples propósitos que cada empresa puede tener a la hora de implementar un Sistema de Gestión del Desempeño en general (y una Herramienta de Evaluación del Desempeño individual en particular). La cantidad de propósitos que se persigan y la prioridad de cada uno de ellos deben estar en línea con la estrategia y los objetivos que a través de ellas se pretendan alcanzar. Cuanto más claros estén los objetivos finales del proceso, tanto más fácil será elegir (o diseñar) la herramienta adecuada y tanto mayores serán las chances de que la inversión de tiempo y dinero dé los frutos esperados.

Entre los propósitos más usuales se incluyen:

- **Transmitir los Valores y la Estrategia de la Organización.**

Desde ya, la primera responsabilidad del *Management* es definir estrategias y valores que le permitan a la organización tener éxito. Una vez definidas la misión y visión de la empresa, qué objetivos se buscan alcanzar, la manera práctica en que cada uno de esos objetivos puede alcanzarse desde cada uno de los roles dentro la organización y explicitados los valores que la empresa busca encarnar, es preciso comunicarlos a todos los departamentos de la compañía y alentar a que los comportamientos y valores sean adoptados por quienes son cotidianamente sus representantes: los empleados.

Una empresa que funciona en línea con su estrategia la tendrá presente al momento de seleccionar a los candidatos, explicitar las expectativas que se tienen para con cada uno, supervisar el cumplimiento de dichas expectativas, asistir a sus recursos a través del entrenamiento e introducir planes de apoyo o mejora cuando fuera preciso. Los objetivos concretos y las metas por alcanzar son elementos que deberían comunicarse y medirse también a través del uso de

evaluaciones de desempeño; es un modo de asegurarse de que la estrategia organizacional está representada adecuadamente por cada uno de sus miembros.

#### - **Alinear Comportamientos y Objetivos**

La performance de cualquier empleado muy rara vez se mantiene estable en el tiempo. Además de los factores personales o motivacionales, la propia dinámica laboral puede ser causa de ello: no es infrecuente que por promociones o reasignaciones de tareas un empleado valioso comience a tener dificultades para mantener un rendimiento satisfactorio. Los motivos pueden ser diversos (por caso, quizás no fueron comunicadas adecuadamente las expectativas, el empleado no tiene todas las capacidades necesarias para desempeñar su nueva posición o sencillamente tiene dificultades para adaptarse a la dinámica de un nuevo grupo de trabajo).

En todos esos escenarios, es preciso detectar la brecha entre el desempeño esperado y el actual a través de la evaluación. Si el empleado sabe qué comportamientos se esperan de él y cuáles son los objetivos que debe cumplir, tras una serie de evaluaciones debería ser sencillo identificar la causa raíz del problema y establecer un modo de proponer soluciones. Esto puede conseguirse más fácilmente si se dispone de una Herramienta de Medición del Desempeño en funciones.

#### - **Definir y Evaluar la Gestión de RRHH**

La gestión de RRHH es la principal responsable de implementar políticas de contratación, remuneración, promociones, desarrollo y gestión que materialicen la estrategia definida por la gerencia para que la empresa tenga éxito. Dicho de otro modo, cumple un rol esencial para que la inversión de esfuerzo, tiempo y dinero acabe teniendo los resultados esperados. Una herramienta que consiga mensurar la performance y el potencial de cada uno de los empleados le aportará al

departamento de RRHH los medios para tomar decisiones a todos los niveles anteriormente mencionados.

Además, tener dicha información histórica y consolidada es especialmente útil cuando se deben tomar decisiones en tiempos de crisis (que demandan reducción de personal) o de crecimiento explosivo (que requiere reorganizar la estructura existente y atraer nuevo talento), ya que ayuda a tener presentes procesos, necesidades y tendencias en un momento en el que el tiempo es escaso y los posibles aciertos y errores tendrán un impacto especialmente alto.

#### - **Mejorar Mecanismos de Contratación**

La definición de lo que se espera que cada empleado produzca es una extensión de las definiciones que la empresa tiene para cada uno de los roles dentro de ella. Como regla general, la selección de personal partirá de una descripción de las habilidades mínimas y deseadas para poder cumplir un rol dentro de la empresa.<sup>2</sup> La evaluación de desempeño (en especial en aquellas herramientas que consolidan *feedback* de múltiples fuentes para caracterizar la performance de un empleado) puede contribuir a redefinir las habilidades buscadas, ya sea porque detectó características salientes en quienes han rendido por encima de las expectativas o porque ha logrado identificar aspectos a mejorar.

Así, analizar la información condensada por las evaluaciones de desempeño puede llevar a redefinir las competencias deseadas. Por caso, agregando capacidades nuevas, matizando la importancia de otras que parecían esenciales, asignándole mayor peso específico a las fortalezas

---

<sup>2</sup> En empresas que dependen de contratar talento escaso y de alto potencial, este modo de contratar está cayendo en desuso. En vez de contratar para llenar puestos vacantes (buscando mantener la performance) se propone contratar talento escaso primero y ver qué puestos pueden asignarse después. Este último método tiene la ventaja de que termina construyendo un pool de talento más heterogéneo que ofrece mejores perspectivas en cuanto a adaptabilidad y flexibilidad. Véase Hatum (2011), pp. 75-78.

que tenía un recurso puntual que se necesita reemplazar o balanceando las habilidades del equipo actual contratando recursos que ofrecen ventajas complementarias.

#### - **Gestionar Ascensos y Decisiones de Remuneración**

Este es uno de los aspectos más extendidos en el uso práctico, y casi siempre un pilar de la política de RRHH.<sup>3</sup> Ascensos, bonus y actualizaciones salariales son mecanismos tradicionalmente usados para motivar y estimular a los empleados, especialmente en mercados muy competitivos donde la lucha para retener el talento es encarnizada.<sup>4</sup> No obstante, también estas prácticas son fuentes de conflicto o daños a las relaciones en el seno de los equipos de trabajo o en el vínculo de éstos con sus líderes. Un modo de minimizar las rispideces es definir objetivos y métricas que serán utilizados para medir el desempeño de los empleados, medirlos con continuidad y reportar los resultados frecuentemente.<sup>5</sup> Con ese respaldo, en teoría, al momento de tomar decisiones de remuneración los equipos tendrán ya presente la posición relativa de cada empleado y aumentarán las chances de que las propuestas sean percibidas como razonables u objetivas.<sup>6</sup>

#### - **Documentar el Rendimiento ante posibles Juicios Laborales**

Un empleado que siente que no se lo ha evaluado objetivamente puede llegar a efectuar reclamos por la vía legal, en especial en países donde existe una alta litigiosidad o la mayoría de

---

<sup>3</sup> Casanova (2016) señala la inconsistencia de ese punto dado que la compensación no es el método más duradero para motivar empleados y alinearlos con la estrategia de la empresa, pero continúa siendo uno de los principales usos.

<sup>4</sup> No obstante, Spolsky (2000) presenta argumentos en contra del uso de pagos únicos por performance incluso (y hasta especialmente) en mercados competitivos.

<sup>5</sup> Un modo más audaz (sobre el que se volverá en el Capítulo IV del presente trabajo) es eliminar las negociaciones y retribuciones individuales y vincularlas solamente a roles, experiencia y capacidad técnica. Para un análisis detallado y una propuesta general sobre esta implementación, véase Spolsky (2009).

<sup>6</sup> No siempre es el caso en la práctica: para empezar, los empleados pueden rechazar la forma de medir las métricas o el modo en que se las ha comunicado. Habría que agregar que las decisiones de remuneración tienen un efecto de muy corto plazo en la motivación de los empleados, y por ende son susceptibles de mejorar la performance sólo por poco tiempo. Como hemos mencionado anteriormente, Casanova (2016) descrea de la utilidad práctica de las Herramientas de Evaluación de Desempeño aún cuando este es el objetivo declarado más extendido. Véase páginas 94-97 para una argumentación más extendida.

la población empleada formalmente trabaja bajo la órbita de una organización gremial. La falta de discriminación, la ecuanimidad de los líderes de los equipos y la ausencia de arbitrariedad en el trato para con los empleados puede respaldarse en la documentación de las evaluaciones de desempeño de todo el equipo de trabajo, en especial cuando consta en ella que el empleado ha podido dar su punto de vista al momento de recibir su evaluación o calificación.

Este uso, lamentablemente, conlleva un gran costo: es responsable de las quejas que suelen acumularse contra la burocratización y la persistencia en utilizar procesos manuales. Son aspectos que una mayoría de los usuarios de estas herramientas en empresas latinoamericanas ven hoy día como una oportunidad de mejora en el proceso.<sup>7</sup>

#### - **Desarrollo de los Empleados y Planes de Carrera**

Este elemento es de aparición relativamente reciente, junto con las Herramientas de Evaluación de Desempeño más modernas. Uno de los mejores aportes que puede devolver el proceso es guiar las decisiones de capacitación de los empleados, en especial si consigue armonizar sus intereses y capacidades con las oportunidades dentro de la empresa. De esta manera, deja de centrar la atención en la brecha entre el rendimiento actual y el deseado para poder enfocarse en la diferencia entre la actualidad y el potencial del empleado, pudiendo establecer una hoja de ruta para el desarrollo.

Cuando el proceso es exitoso se materializa una sinergia entre las necesidades y oportunidades de la empresa y el interés del empleado, convirtiendo al *feedback* en motivador y desafiante. Como es lógico, este enfoque es mucho más fácil de robustecer a través de

---

<sup>7</sup> En Focaia y Moya (2018) se documenta que el 65.38% de los usuarios de Herramientas de Medición de Desempeño no consideran al sistema amigable, ágil ni dinámico (adaptable) mientras que el 73.40% cree que la evaluación de desempeño no acaba por integrarse con otros procesos de RRHH.

Herramientas de Evaluación de Desempeño que no tienen procesos estructurados y hacen participar activamente al evaluado y su equipo de trabajo.

- **Maximizar Beneficios de la Empresa**

Este último propósito se desprende del supuesto de que una creciente productividad de cada uno de los empleados redundará necesariamente en mayores beneficios para la empresa. Una advertencia importante debe matizar esta aparente tautología: una Herramienta de Evaluación de Desempeño mal diseñada o implementada probablemente genere el efecto contrario, dado que no solamente puede afectar negativamente la dinámica de los grupos de trabajo o desmotivar a empleados valiosos, sino que para peor exige dedicar recursos, tiempo y dinero. Quizás por ese motivo, existen empresas que tras implementar fallidamente un esquema de evaluación de desempeño optan por abandonarlo y hasta pueden argumentar que dicho rechazo ha aumentado su rentabilidad.<sup>8</sup>

## **I.B. Métodos de Evaluación**

En esta sección se hace un breve repaso de los métodos de evaluación más conocidos. Dado que se mencionan brevemente las limitaciones de cada uno, se opta por comenzar por los más antiguos, con la esperanza de que la evolución entre esos métodos y los más modernos sea más fácil de apreciar.

- **Método Narrativo**

Es esencialmente una lista de los aciertos y fallas que el trabajo de un empleado ha presentado a los ojos de su evaluador. Si la calidad del análisis es alta y el evaluador consigue ser

---

<sup>8</sup> Encontraremos casos representativos en el Cuerpo Empírico del presente trabajo.

exhaustivo y objetivo, el empleado tendría la posibilidad de comprender sus fortalezas y debilidades, además de poder enfocarse en qué debe mejorar.

Este método, como es lógico, depende enormemente de la capacidad y objetividad del evaluador, además de padecer la limitante de no permitir la comparación entre individuos. Por ese motivo, no es apto para satisfacer muchos de los objetivos presentados en el apartado anterior. No obstante, correctamente aplicado tiene la ventaja de ser simple y veloz.

#### - **Ranking de Empleados**

A diferencia del anterior, es un método relativo: consiste en el ordenamiento de la performance de los empleados del mejor al peor. Como es lógico, sólo es relevante utilizarlo cuando el tamaño del grupo a evaluar es suficientemente grande. Debe, además, tener sentido práctico comparar los distintos trabajos individuales entre sí: por ejemplo, sería difícil de implementar en un equipo compuesto por 4 programadores, 2 analistas de control de calidad, 1 analista funcional y 1 gerente de proyecto, dado que los objetivos de cada recurso no serán uniformes con los de la mayoría de sus compañeros.

Una desventaja inicial es que el ranking estará afectado por la subjetividad del evaluador. Además, por su naturaleza este método no está enfocado en los procesos que determinan las diferencias de rendimiento ni en cómo contribuir a que mejoren los de bajo rendimiento. Por último, quienes ocupen los últimos lugares del ranking probablemente se desmotiven o comiencen a limitar su colaboración con el equipo.

#### - **Comparación de a Pares**

Comparte con el método anterior el ser un método de ordenamiento relativo. A diferencia de éste, se hace confeccionando una matriz de doble entrada donde cada empleado es comparado

con todos los demás. Las dimensiones a lo largo de las cuales se comparará a los empleados deben estar definidas de antemano, y ser representativas de un objetivo estratégico para la organización. Una vez completado el ejercicio, se ordena al conjunto de los empleados según la cantidad de veces que cada uno de ellos ha sido el mejor en una comparación contra un compañero.

La principal desventaja de este método (además de las que comparte contra el ordenamiento simple, es decir, su falta de foco en procesos y mejoras y su posible arbitrariedad) es que insume mucho tiempo por parte del evaluador, tanto realizando la comparación como obteniendo la información que operacionalizará las dimensiones para que sirvan al ejercicio.

#### - **Distribuciones Forzadas**

Se trata de un tercer método relativo, que simplifica muchísimo la dinámica respecto de los anteriores. Por ejemplo, se predefine una proporción determinada de empleados que rendirán de forma extraordinaria, otra que se desempeñará satisfactoriamente y una tercera que presentará rendimientos insuficientes.<sup>9</sup> Luego, las evaluaciones de desempeño buscarán medir el rendimiento de cada empleado y proveer la información que asista al evaluador para ubicar a cada empleado de modo tal que la población se distribuya según la proporción conocida de antemano. Típicamente, la gran mayoría de los empleados se ubicará dentro del rendimiento satisfactorio.

Como se intuye, es un método simple que además de ser relativamente rápido obliga a los evaluadores a eludir la solución de compromiso de evaluar a los empleados con la misma calificación. Por otra parte, facilita las retribuciones variables o los ajustes de desempeño, que

---

<sup>9</sup> El tamaño de cada sección e incluso la cantidad de secciones varía según los intereses de la empresa. Durante las entrevistas que sirvieron de fuente al Cuerpo Empírico del presente trabajo se escuchó el caso de una empresa que armó 4 categorías, donde la distribución (de peor a mejor calificados) es 10%-30%-40%-20%. Véase Hatum (2009), pp. 116.

pueden ser predefinidos de antemano ya que la distribución de la población es conocida incluso antes de que se completen las evaluaciones.

No obstante, el método está lejos de ser la panacea: para empezar, suele conspirar contra la colaboración dentro de los equipos y no está exento de los resquemores que siempre se dan a la hora de elaborar un ranking relativo de empleados. Además, en equipos relativamente estables en el tiempo, el evaluador tiene la posibilidad de ir rotando las calificaciones de sus evaluados, de modo tal que a mediano plazo todos acaben por recibir la misma cantidad de calificaciones de cada tipo. El evaluador minimizaría de ese modo las rispideces que le implica la distribución forzada, pero vaciaría subrepticamente de sentido a toda la herramienta.

#### - **Evaluación de Comportamientos**

Se trata de representar el grado en el cual un empleado presenta una serie de comportamientos, generalmente medidos en una escala de tipo Likert (típicamente de entre 4 y 5 grados posibles).<sup>10</sup> La gerencia o el departamento de RRHH es responsable de definir categorías y la lista de comportamientos individuales, lo cual debería ocurrir teniendo en cuenta la estrategia de la empresa. Cada familia de comportamientos suele recibir una ponderación, con el fin de que el resultado final de la evaluación sea representativo de la forma en que el empleado se ha desempeñado en las que tienen más importancia o son más fomentadas. Una vez que está terminada la lista de comportamientos a relevar, el evaluador debe calificar a cada empleado. Para que con esta información sea viable confeccionar un ranking de todos los empleados, se debe finalmente

---

<sup>10</sup> Suelen asignarse valores de 1 (para el menor acatamiento) a 5 (para el mayor). También pueden representarse como una frecuencia (donde las escalas irán de “nunca” a “siempre”) o nivel de aceptación de un postulado (“nada de acuerdo” – “totalmente de acuerdo”). Independientemente de ello, las respuestas pueden luego traducirse para ser procesadas y comparadas de forma cuantitativa.

computar en cada evaluación individual un promedio ponderado de observancia de los comportamientos.

Esta herramienta tiene la gran ventaja de incorporar aspectos cualitativos y luego operacionalizarlos, con lo cual señala (al estilo del método narrativo) áreas donde cada empleado se destaca y cuáles deben mejorarse sin sacrificar la posibilidad de ordenar a los empleados según rendimiento. Lamentablemente, el hecho de que los evaluadores no puedan participar de la confección de la herramienta y que la misma deba usarse para toda la empresa puede terminar viciando las evaluaciones cuando la cantidad de empleados es muy grande y los objetivos de cada rol son suficientemente diversos.

#### - **Métodos Gráficos**

Independientemente de su formato y el modo en que son representados, de lo que se trata es de presentar una lista de factores a evaluar y predefinir niveles de performance para cada uno de ellos. Para cada tópico a evaluar, dichos niveles de performance se detallarán con ejemplos de comportamientos que ilustrarán cada una de las calificaciones posibles. De tal modo, el ejercicio de la evaluación se simplifica porque el evaluador sólo debe detectar la presencia de ciertos comportamientos observables y limitarse a presentárselos al evaluado.

Este tipo de métodos cuenta con muchas ventajas, lo cual explica su popularidad: la evaluación es veloz, los criterios que definen la performance de los empleados son fáciles de entender por parte de los evaluados y son modos eficaces para traducir la estrategia de la empresa en comportamientos observables. Al igual que las evaluaciones de comportamientos, sin embargo, presenta limitaciones. Primeramente, el diseño de la herramienta demanda mucho tiempo y esfuerzo. Además, si la organización fuera de tamaño grande o tuviera sectores o rubros de

objetivos y prácticas muy diferentes (por ejemplo, el sector de finanzas y el sector de soporte técnico) la herramienta deberá elegir entre ponderar los comportamientos según sector (sacrificando el ordenamiento de los empleados a nivel empresa) o mantener los comportamientos de modo uniforme (corriendo el riesgo de que la evaluación no sea representativa del verdadero aporte de cada empleado al cumplimiento de los objetivos de su sector).

#### - **Objetivos de Desempeño (MBO, del inglés Management By Objectives)**

Es un método capaz de superar varias de las limitaciones del método anterior. Operacionaliza los objetivos de corto, mediano y largo plazo de la estrategia de la empresa solicitándole a los gerentes que los conviertan en objetivos prácticos para sus líderes de equipo, que a su vez deberán expresarlos como objetivos que traduzcan el aporte que se espera de cada empleado. Si la implementación es exitosa, entonces, los objetivos individuales terminarán por contribuir a los de la empresa. Junto con los objetivos, se definen también las métricas que servirán para medir el cumplimiento o no de dichos objetivos. Buena parte del éxito del método pasa por el correcto diseño de las métricas, la viabilidad de medirlas con frecuencia en plazos razonables y (como es evidente) la destreza y capacidad de gerentes y líderes para traducir los objetivos corporativos en otros atómicos y hechos a la medida de los miembros de su equipo.

Si bien en la dinámica propuesta hasta aquí los objetivos se desprenden formalmente de la estrategia de la empresa, es también posible adoptar un enfoque levemente distinto. En esta alternativa, cada gerente o líder confecciona objetivos que combinan las necesidades de su área con la estrategia de la empresa, o incluso objetivos innegociables del área cuyo vínculo con el éxito estratégico de la empresa no es directo. La combinación de estos aspectos resulta en objetivos híbridos; su relación con la estrategia corporativa adopta formas no siempre evidentes, cuya

complejidad o independencia aumenta conforme la implementación baja de nivel en el escalafón de la empresa. En estructuras corporativas donde cada nivel crea sus objetivos híbridos sólo cuando ha terminado de hacerlo el nivel jerárquico superior, estas modificaciones se hacen en secuencia, como si los objetivos fluyeran de arriba abajo como el agua de una cascada. Por este motivo, a esta forma híbrida de implementación se la suele conocer (echando mano a un neologismo) con el nombre de objetivos que *cascadean* informalmente de la estrategia.

Aunque es un método relativamente extendido, es también de los que requieren más esfuerzo y menos facilita la comparación entre empleados de distintos sectores de la empresa. También requiere más ciclos de prueba y error hasta que finalmente alcanza velocidad crucero. No obstante, cuando la implementación resulta exitosa, es de los que más réditos le trae a la empresa.

#### - **Revisión del Trabajo**

Puede entenderse como una fracción del método anterior. Busca medir los objetivos atómicos que se desprenden de la descripción de cada puesto de trabajo, apelando tanto a la opinión del evaluador como la información provista por pares, clientes internos y clientes externos. El propósito es que la evaluación le permita al empleado comprender qué debe hacer para mejorar su rendimiento, incluyendo soporte, capacitación y entrenamiento.

Este método, por su aporte cualitativo, suele ser parte complementaria de Herramientas de Gestión del Desempeño más rígidas, como el ordenamiento de empleados o la distribución forzada. Por sí solo, tiene como méritos la consolidación de información de diferentes fuentes y la rápida posibilidad de acuerdos entre evaluador y evaluado. No obstante, sin complementarlo con otras herramientas no permite comparación con los pares, y tampoco garantiza que los objetivos definidos por el evaluador contribuyan eficazmente con la estrategia de la empresa.

### - **Evaluación por Objetivos con Participación del Evaluado**

Este método se asienta sobre los mismos principios que el método MBO mencionado anteriormente, con la enorme diferencia de que el evaluado participa en el diseño de la herramienta desde el punto de partida. En esencia, evaluador y evaluado consensuarán objetivos, modo de alcanzarlos y métricas que permitirán validar los avances. Pasado ese punto, el empleado asume la responsabilidad de sostener el plan acordado, mientras que el rol del evaluador pasa a ser el de proveer *feedback*, asistirlo en la redefinición de planes, proveer sugerencias y en general brindar apoyo durante todo el proceso.

Esta variante del MBO tiene la ventaja de que involucra más a cada empleado y es más susceptible de motivarlo. Por lo demás, de ser implementada con eficacia, sus virtudes y limitaciones son similares a las del MBO clásico referido anteriormente.

### - **Autoevaluación**

En este método la participación del evaluador se suaviza aún más. Su responsabilidad fundamental es proveerle al evaluado un conjunto de categorías que deberían estar vinculadas con los objetivos y valores de la empresa, capacitándolo para poder mensurarlas y generalmente proveyendo un formulario que sirva de guía. El evaluado debe entonces buscar hechos donde anclar su percepción del trabajo que ha realizado, condensarla en el soporte que le han provisto y cerrar un ciclo del proceso en una entrevista con su evaluador, durante la cual se suele arribar a una calificación que resuma el proceso.

Más allá de la alta dependencia del esfuerzo del evaluado, el producto de aplicar este método es común a varios de los anteriores: a través del autoanálisis, el empleado debería poder conocer sus fortalezas, sus áreas de mejora y tener una mejor perspectiva de lo que necesita para

continuar su desarrollo. El principal aporte del método no está tanto en resumir el rendimiento pasado o *rankear* al empleado respecto de sus pares, sino en incentivarlo a que su rendimiento futuro se acerque a su potencial. Pese a ello, o quizás precisamente por este motivo, no es infrecuente que este método se combine con otros como Evaluaciones de Comportamientos y MBO. A cambio del esfuerzo adicional invertido en esas hibridaciones, se busca tanto obtener las ventajas de ambas herramientas como tener un método alternativo de evaluación para el caso de que la implementación de una de ellas no arroje los resultados esperados para determinados casos puntuales.

#### - **Evaluación 360°**

Se trata de la más completa evaluación posible sobre la performance de cada empleado, en tanto no solamente incluye a sus líderes y la autoevaluación, sino que busca consolidar *feedback* de todos los actores que interactúan con el evaluado, incluyendo clientes internos, clientes externos, proveedores, pares y subordinados. Todos los participantes responden una serie de preguntas que deberían arrojar luz sobre fortalezas y debilidades del evaluado y (como es deseable que ocurra en toda evaluación) necesitan apoyarse en ejemplos prácticos y antecedentes para anclar las impresiones en elementos reales. Luego, todas las respuestas se consolidan en un reporte específico para cada empleado, que no solamente deberá subrayar las fortalezas y debilidades: pondrá en el candelero especialmente a los elementos donde la percepción del evaluado no coincide con la de su entorno, con el fin de trabajar sobre ellas. Por este motivo, la herramienta es especialmente útil para diseñar planes de desarrollo, a tal punto que ese era su propósito original.

Si bien la digitalización de las últimas décadas contribuyó a acelerar el procesamiento y la consolidación de este tipo de ejercicios (que era increíblemente farragosa en papel), la cantidad de

información recopilada y su análisis inevitablemente demandan todavía un gran esfuerzo de toda la compañía. Esto ocurre especialmente en equipos numerosos, donde todos los evaluados deberán esmerarse para dar un *feedback* enriquecedor para con todos sus compañeros además de hacer sus propios ejercicios introspectivos. El volumen de esfuerzo conspira contra la posibilidad de generar nuevas revisiones con la periodicidad necesaria, con lo cual el ciclo de evaluación y desarrollo de los empleados suele terminar cortándose antes de lo ideal.<sup>11</sup>

#### - **Evaluación basada en procesos no estructurados**

Este método nace como forma de combatir aspectos burocráticos y concentrarse en la transparencia y frecuencia de los encuentros. Sacrifica toda posibilidad de comparación entre los empleados y quita la mayor cantidad de burocracia posible: lo único esencial es que el par evaluador/evaluado se reúna con frecuencia periódica para conversar (partiendo de hechos observados que obren como disparadores) sobre qué aspectos del trabajo deben mantenerse, cuáles deben hacerse de otro modo y cuáles deben interrumpirse.<sup>12</sup>

Implementado a conciencia, el método logra sostener un *feedback* rápido y frecuente y generar un vínculo más fluido entre evaluador y evaluado, lo cual suele traer aparejado un mayor compromiso del empleado con su propio desarrollo y una más fluida introspección sobre sus virtudes y posibles áreas de mejora. Como contrapartida, desde su propia concepción sacrificará

---

<sup>11</sup> La velocidad y la eficiencia son grandes pendientes para este tipo de herramienta. No obstante, cumple con muchos de los nuevos principios para los Sistemas de Gestión de Desempeño: la gente como protagonista, el protagonismo de los pares, la diversidad y el foco en la eliminación de los prejuicios, enfocarse en el desarrollo y el bienestar, darle centralidad a la transparencia y a las posibilidades ofrecidas por el mundo digital. Véase Molinari (2017).

<sup>12</sup> Una versión grupal de este método es práctica habitual en las reuniones retrospectivas que cierran ciclos de Metodologías Ágiles de desarrollo de software, lo cual explica que ciertos autores recomienden a equipos familiarizados con el ejercicio que lo repliquen a nivel individual. Véase Pierce (2011).

la posibilidad de consolidar grandes cantidades de información y comparar el rendimiento entre evaluados.

### **I.C. Participantes del Proceso**

Durante la caracterización de las más familiares Herramientas de Evaluación de Desempeño, se ha mencionado a los participantes del proceso. El rol (o incluso los diversos roles) que le toquen a cada miembro de la empresa dependerán del modo en que se definan el Sistema de Gestión de Desempeño y la dinámica derivada de los Métodos o Herramientas de Medición elegidos.

Los cuatro grandes grupos de participantes de un proceso de evaluación de desempeño son la gerencia, los evaluadores, los evaluados y terceros interesados en la existencia o los resultados del ejercicio (dentro o fuera de la empresa). Se mencionará a continuación las responsabilidades y experiencias de cada grupo.

#### **- Gerencia**

Este grupo es responsable de las decisiones clave que incidirán en el éxito de la Herramienta de Evaluación del Desempeño. Deben traducir la estrategia de la empresa en una serie de objetivos a medir, elegir el sistema más adecuado para conseguirlo, convertirlo en una prioridad dentro de la empresa y frecuentemente autorizar la inversión, ya sea en tiempo, costo o materiales.

Es vital que este encadenamiento de objetivos y los modos a través de los cuales alcanzarlos guarden coherencia interna. Cualquier ruptura entre estos eslabones hará que los resultados disten de los esperados. Por caso, si los managers eligen una herramienta cuya principal ventaja es la velocidad para establecer relaciones entre el desempeño y decisiones de remuneración, el objetivo último no puede ser el desarrollo de los empleados o potenciar su motivación. Otro ejemplo, desde

la perspectiva de los costos, sería implementar un sistema muy demandante en tiempo y recursos cuando no está asegurado que el presupuesto pueda mantenerse en el tiempo. Una visión coherente por parte del Management, por lo tanto, es condición indispensable para que los demás participantes del proyecto tengan chances de llevarlo a buen puerto.

#### - **Evaluadores**

Una vez seleccionados los métodos para evaluar el desempeño, los evaluadores tendrán la responsabilidad de llevar a la práctica el proceso sin desnaturalizarlo. Su tarea demanda esfuerzo, tiempo y dedicación: deberán entender cómo funciona la herramienta, ser metódicos en las observaciones para con todos sus evaluados, ejercitar todas las destrezas necesarias para la presentación en las reuniones de *feedback* (en especial el modo de presentar críticas constructivas para motivar a sus evaluados) y entrenarse para manejar los efectos que pudiera tener el proceso en su relación con los evaluados.

Tener responsabilidad sobre las oportunidades de desarrollo o promociones de los evaluados les implica ya un desafío en lo personal y lo emocional, pero los evaluadores deben absorber presiones incluso mayores cuando el producto del proceso pueda relacionarse con decisiones de remuneración o (en casos extremos) con desvinculaciones de la empresa. Por esa razón, es necesario que los evaluadores reciban tanto el entrenamiento necesario para afrontar estas situaciones de modo profesional como los incentivos adecuados por parte del Management. Sin esos incentivos para que dé *feedback* frecuente, puede ocurrir que las demás responsabilidades de gestión por parte del líder terminen quitándole el tiempo mínimo para desarrollar esta tarea.

Así como se subraya que un evaluador insuficientemente calificado o poco comprometido con la tarea puede hacer naufragar una implementación que de otro modo hubiera sido exitosa,

debe admitirse que ni siquiera los evaluadores más aplicados tienen posibilidades de mejorar un sistema de evaluación inadecuado para los objetivos. Hemos ejemplificado esta limitación líneas atrás haciendo referencia al recurso de rotar calificaciones entre los empleados para atenuar el impacto negativo de las distribuciones forzadas en equipos donde debiera más bien priorizarse el desarrollo. Aunque a esas limitaciones solamente se les puede oponer tácticas estériles y de vida útil muy corta, son a veces las únicas alternativas a mano del evaluador para atenuar el impacto de una incorrecta decisión original de la gerencia.

#### - **Evaluatedos**

Reverso de los participantes presentados previamente, componen generalmente a la mayoría de los miembros de la empresa. Abarcan distintos roles, responsabilidades, niveles de experiencia y grados de motivación. Del involucramiento que tengan con el método y su motivación para responder a él dependerá en última instancia que se cierre la brecha entre los objetivos de la empresa y su realidad presente.

El evaluado es consciente de las expectativas que la empresa tiene de él al momento de hacer una evaluación, pero esto no quita que no tenga a su vez expectativas propias para con el proceso. Como el resultado de una evaluación es generalmente consolidado en su perfil dentro de la empresa, hay consciencia de que incidirá en sus futuras oportunidades profesionales, su retribución fija y (si aplicara) en su retribución variable.

Para que el proceso no dañe su vínculo con el evaluador y termine por alienarlo de la cultura de la empresa, el evaluado debe percibir que durante el proceso su rendimiento ha sido analizado en detalle, con seriedad, respeto y ecuanimidad. Se le debe haber dado un espacio donde se escuchó su propia opinión sobre el proceso y el resultado. Para que la evaluación sea percibida como justa,

además, las expectativas sobre las que se lo evalúa le tienen que haber sido comunicadas con suficiente antelación y deben estar vinculadas con objetivos relevantes y alcanzables. En aquellos casos donde los resultados de la evaluación incluyen un componente relativo, tienen que haber elementos para que el evaluado se sienta justamente comparado con otros miembros de su equipo, sin que hayan mediado favoritismos o elementos ajenos a los relacionados con el método de evaluación. Además, el evaluador tiene que actuar como un interesado en el desarrollo del evaluado, en lugar de adoptar la postura de un juez que le señale aciertos y errores.

En resumen, el evaluado debe sentirse respetado, valorado e incluido a lo largo del proceso para que un Sistema de Gestión del Desempeño traiga resultados positivos y no dañe más que lo que aporta.

#### - **Terceros Interesados**

Estos protagonistas pueden estar dentro o fuera de la empresa en cuestión. Son todos los actores que se nutren de los procesos del Sistema de Gestión del Desempeño y la información que ellos consolidan a través de las Herramientas. Entre ellos se encuentran sectores de *Compliance* de la empresa (que están atentos a que los resultados de las evaluaciones no estén afectados por estímulos o acuerdos irregulares), tribunales del fuero laboral (que según la eficacia del sistema pueden ver en el corpus de las evaluaciones de desempeño un trato desigual para con ciertos empleados o la cristalización de reglas claras para todos los miembros de la empresa), sindicatos (interesados en lo que refiere a negociaciones salariales en las que un empleado se siente perjudicado o en evaluaciones posiblemente injustas que han derivado en un despido) y profesionales o académicos (interesados en el aprendizaje que pueda proveerles el resultado práctico de cada implementación).

Aunque son los protagonistas menos directos si centramos nuestra atención en las decisiones de una empresa particular, sus intereses y motivaciones tienen incidencia clara: obran simultáneamente como proveedores de información valiosa para enriquecer a los Sistemas de Gestión de Desempeño, custodios de la presencia de un conjunto mínimo de transparencia, supervisores de la existencia de reglas y delimitadores del campo en el cual la empresa puede implementar su sistema.

#### **I.D. Características, Desafíos, Recomendaciones y Limitaciones**

Los Sistemas de Gestión de Desempeño, incluso hoy día, continúan teniendo en Latinoamérica tanto una alta utilización como un bajísimo porcentaje de aprobación.<sup>13</sup> Este hecho sugiere que la mejora de rendimiento prometida por la implementación de estos sistemas está lejos de ser automática. Las principales amenazas que enfrenta una implementación son la ausencia de características prácticas mínimas, la insuficiente capacitación de los participantes, las fallas al momento de dar *feedback* o los condicionamientos por parte de terceros interesados que desnaturalicen el proceso.

En lo que refiere a las características fundamentales que debe tener un Sistema de Gestión del Desempeño, se pueden destacar las siguientes: Relevancia (debe capturar los objetivos para que cada rol contribuya al éxito de la empresa y luego ser capaz de medirlos a lo largo del tiempo), Identificación (debe poder distinguir entre buena y mala performance con cierta objetividad), Confiabilidad (los resultados de la evaluación no deberían variar significativamente si diferentes evaluadores debieran analizar la misma información), Practicidad (debe ser simple de implementar, comprender y aplicar y no insumir más tiempo o recursos que los previstos) y

---

<sup>13</sup> Focaia y Moya (2018) presentan resultados por los cuales sólo el 40,71% de empresas Latinoamericanas evalúa satisfactoriamente los resultados de sus Sistemas, pese a que el 78.53% de los relevados cuenta con un sistema formal.

Aceptabilidad (el Management debe generar y mantener el involucramiento activo de todas las partes).

Si cada una de estas características suena evidente, la frecuente insatisfacción con los resultados desmiente su trivialidad. Es concebible que un programa bien diseñado se desalinee con los propósitos de la empresa en caso de un cambio estratégico o de rubro. Por ejemplo, una empresa que migra de proveer servicios a vender productos deberá repensar buena parte de su Sistema, y no es inhabitual que dichas modificaciones se introduzcan con posterioridad al cambio de roles y reorganización de tareas. Tampoco es necesariamente simple que se puedan identificar con confiabilidad performances inadecuadas, tanto por el hecho de que nunca puede eliminarse la subjetividad de los evaluadores como por la propia naturaleza de algunos de los métodos. Por último, tanto la practicidad como la aceptabilidad (asumiendo que tuvieran un exitoso punto de partida) bien pueden degradarse conforme la compañía varía sustancialmente en cantidad de empleados o enfrenta nuevos desafíos. En suma: es infinitamente más sencillo definir las características mínimas deseables que debe tener un Sistema de Gestión que garantizarlas en la práctica.

En lo que respecta a la capacitación de los participantes, la empresa debe ante todo presentar a todos los evaluadores las razones por las cuales se implementa el programa, qué características tiene, qué se espera de ellos desde un punto de vista operativo, cómo se espera que consolide su *feedback* en modo de información o calificaciones y cuáles son los errores típicos junto con el modo más tradicional de evitarlos. Las habilidades que se esperan del evaluador son diversas y complejas independientemente del Método que éste siga, pero incluso evaluadores competentes y experimentados pueden rendir por debajo de sus capacidades frente a un método para el que no han sido específicamente entrenados.

También es preciso entrenar a los evaluadores para que atenúen el impacto subjetivo de las evaluaciones. Las capacitaciones buscarán preparar al evaluador para escenarios usuales. Pueden apoyarse en la observación de casos donde se cometen errores típicos o las propuestas sobre qué hacer en instancias donde las calificaciones caen en zonas límite. A través de ese ejercicio se buscará atenuar la intensidad de los sesgos más comunes, entre los que se cuentan: la excesiva severidad o generosidad de un evaluador puntual para con su equipo, la miopía que le hace a un evaluador basar la evaluación en lo que recuerda (especialmente del pasado reciente en vez del conjunto de los comportamientos que se desarrollaron en el período), la comparación involuntaria entre rendimientos de evaluados cuando el método debería enfocarse nada más que en el desarrollo de cada uno por separado, la afinidad que lo lleva a reconocer el rendimiento de un empleado inconscientemente por motivos personales o el “efecto halo”, que sobrevalora los comportamientos o el desempeño de un evaluado debido a una impresión positiva preexistente (por ejemplo, de períodos de trabajo anteriormente analizados).

Las sesiones de *feedback* son determinantes para el éxito de la implementación del sistema. El modo en que se comunique el contenido de la evaluación incidirá enormemente en la motivación del evaluado, en su comprensión sobre las áreas de mejora y por lo tanto en su capacidad de superar limitaciones actuales en el futuro cercano. Es recomendable que las empresas preparen a sus evaluadores para que estén capacitados para comunicar positivamente, ejercer la escucha activa, comunicar críticas de modo positivo, retomar las riendas en las conversaciones difíciles y dirigir las a un ámbito productivo, mantener el foco en procesos y objetivos más que en la calificación de hechos pasados y encontrar el mejor modo de plantear conversaciones para redefinir metas u objetivos cuando fuera pertinente.

Del mismo modo, se propone que el evaluador incentive al evaluado a que se prepare para las sesiones de *feedback* llevando adelante ejercicios introspectivos y consolidando el producto de éstos en una autoevaluación. A través de ese proceso, es mucho más factible que la instancia de evaluación se convierta en un diálogo apoyado en el entendimiento común, lo cual permitirá tanto al evaluado sacar el máximo beneficio del ejercicio como al evaluador poder entender las causas del rendimiento en un contexto más amplio. También será más sencillo proponer soluciones a situaciones problemáticas que el evaluado pueda haber identificado como una limitante para su performance o desarrollo.

Por último, las empresas solamente pueden adaptar sus Herramientas de Medición del Desempeño hasta cierto punto. Hemos mencionado que hay terceros cuyos roles condicionan el funcionamiento de los Sistemas, entre los que se contaban aquellos encargados de velar por la ley laboral. Mientras persiguen el objetivo de prevenir abusos y arbitrariedades, estos participantes le presentan exigencias al proceso. Entre otros, que genere documentación que describa exhaustivamente los puestos de trabajo, explique cómo dichos objetivos pueden ser medidos por los métodos elegidos, certifique que los resultados han sido comunicados fehacientemente y han dado lugar a réplica donde fuera preciso, demuestre que la capacitación de los evaluadores y la producción de las calificaciones siguen criterios uniformes y (especialmente en caso de desvinculaciones) que la empresa ha hecho lo que estaba a su alcance para ayudar a mejorar al empleado que no estaba rindiendo en el nivel esperado.

Por supuesto, incluso si todos estos aspectos son atendidos en un Sistema de Gestión del Desempeño, no hay garantías de que la propia falibilidad de los procesos o subjetividad de los evaluadores no abra la puerta a reclamos legales. Al fin y al cabo, la interacción humana dejará siempre un margen para la interpretación. Lo que se busca destacar a través de la mención de estas

limitaciones es que en ocasiones la empresa no tiene un modo viable de quitar la resistida burocracia de los procesos: la documentación es a veces un requisito para intentar minimizar los costos legales. En casos extremos, puede que se deba decidir entre afrontar esos costos a cambio de maximizar la aceptación del Sistema de Gestión y mantener la motivación de los participantes. Esto último es lo que hacen las empresas que abandonan sus Sistemas de Gestión del Desempeño y los reemplazan por un mecanismo informal.

La ausencia de una Herramienta de Evaluación no significa que no exista ningún grado de evaluación sobre el trabajo realizado. En empresas donde los equipos trabajan en redes o cada individuo debe desempeñar múltiples roles, las prácticas del grupo deben alinearse con sus objetivos, y el equipo de trabajo intuitivamente logra ponderar los aportes de cada miembro o identificar a los más expertos en cada especialidad. Del mismo modo, la Gestión del Desempeño puede continuar existiendo si cada empleado dedica muchas horas de su tiempo a dar y recibir *feedback* de su equipo de trabajo y de sus líderes directos e indirectos; nadie estará más capacitado que quien tiene la responsabilidad del trabajo operativo para identificar qué tareas le resultan simples y las maneja con naturalidad, y en qué áreas necesita desarrollar nuevas habilidades para poder estar a la altura.

La responsabilidad principal de los gerentes y líderes en ese contexto es correrse del centro de la escena para convertirse en facilitadores del desarrollo de los empleados y en proveedores de oportunidades para que alcancen su potencial. Cuando logran ejercer esa función con eficacia, es perfectamente posible que los beneficios obtenidos del control de los evaluados sobre sus tareas cotidianas, la motivación que les darán nuevas oportunidades y el desarrollo de su potencial atendiendo a necesidades prácticas terminen superando los posibles costos de abandonar un sistema formal.

Se mantendrán presentes estos enfoques informales mientras se caracteriza en el siguiente capítulo a las Metodologías Ágiles para Desarrollo de Software.

## Capítulo II: Desarrollo de Software Siguiendo Metodologías Ágiles

Las Metodologías Ágiles han irrumpido en la industria de software global durante los últimos veinticinco años, y los equipos basados en la Argentina no han sido la excepción. Dichas metodologías se caracterizan por reemplazar el antiguo método “de cascada” (donde los requerimientos se analizaban y consolidaban antes de que el equipo técnico fuera involucrado en el proyecto) por un modelo donde el desarrollo es iterativo e incremental. Se las llama “Ágiles” por contraposición a los métodos tradicionales, que a través de sus reglas y procesos definidos se comenzaron a percibir como pesados y burocráticos.<sup>14</sup>

Precisamente por esas características, estas metodologías enfatizan habilidades que no eran comúnmente centrales en el negocio de desarrollo de software, entre las que se cuentan la colaboración, el trabajo en equipo, la disposición para hacer *coaching*, la capacidad de transmitir ideas y el esfuerzo orientado a compartir conocimiento. De tal modo, la valoración de los miembros del equipo intuitivamente incluye el énfasis en el pensamiento creativo, la capacidad de organizarse y las habilidades sociales. Esto sugiere que la forma de medir el desempeño debería seguir lineamientos compatibles con esa intuición.

En el presente capítulo se caracteriza la dinámica de las Metodologías Ágiles de Desarrollo de Software, así como también los principios de las cuales se desprenden y las ventajas relativas que ofrecen. Además, se menciona qué métricas grupales e individuales se usan frecuentemente para medir avances y grados de éxito. Finalmente, se da cuenta de las dificultades habituales que

---

<sup>14</sup> Para una mayor caracterización de las metodologías ágiles, los principios que las rigen y sus ventajas sobre otras, ver Gonçalves (2019).

suelen presentarse cuando se intenta medir el desempeño a través de Métricas (o las herramientas descritas en el Capítulo 1), y de qué modo se intenta superarlas.

## II.A. Dinámica General

En las Metodologías Ágiles, los requerimientos son capturados desde la perspectiva de los clientes, en vez de ser consolidados por un analista funcional. Dichos requerimientos son recopilados en *User Stories* (o “historias del usuario”), frecuentemente expresadas en lenguaje coloquial del siguiente modo: “Como <rol del usuario>, quiero <funcionalidad deseada> para lograr <propósito>”.<sup>15</sup> El equipo de desarrollo consolida el detalle de las expectativas del cliente en “criterios de aceptación” específicos para cada *User Story*. Para que una *User Story* esté terminada, el criterio de aceptación debe haber sido cumplimentado a criterio del equipo de desarrollo (incluyendo tanto a los programadores como al control de calidad). Las *User Stories* suelen agruparse por *Features* (es decir, se las vinculará según área o nueva funcionalidad que busquen implementar). Completar una o varias *Features* suele ser parte de un mojón en el proceso de desarrollo, a veces llamado *Epic* o *Milestone*.

Para poder estimar el esfuerzo y la complejidad del proyecto, a cada una de dichas *User Stories* se la suele descomponer en *Tasks* (“tareas técnicas”) que deben a su vez completarse para que la *User Story* esté finalizada. Dichas *Tasks* (que pueden también fraccionarse en *sub-tasks* cuando el esfuerzo requerido es demasiado grande o complejo) son creadas por el equipo técnico que estará encargado de la implementación.

---

<sup>15</sup> Por ejemplo, “Como usuario registrado, quiero poder acceder al carrito de compras para conocer el costo total de un pedido antes de pagarlo”. Hay equipos de trabajo que agrupan *Stories* interrelacionadas en elementos que los contienen, a los que se suelen llamar *Epic*. Siguiendo el ejemplo, “Implementación de un Carrito de Compras” puede ser el nombre de una *Epic*.

A cada *Story* y sus subcomponentes se le suele asignar una complejidad, un esfuerzo estimado y (si fuera necesario) detalles técnicos de la implementación para que estén al alcance de todo el equipo de desarrollo. La clave para que el esfuerzo sea razonablemente estimado pasa por dividir el trabajo en elementos cuya complejidad no sea demasiado grande ni demasiado pequeña. Como ventaja adicional, un tamaño intermedio permitirá ver más fácilmente avances cotidianos.

Es importante subrayar que la complejidad se ha de mensurar en una unidad de medida que no tenga relación con el tiempo. A través de su experiencia y entendiendo la solicitud del cliente, el equipo técnico tiene elementos para saber cuán sencilla o compleja es una solicitud: conocen los riesgos, si lo requerido puede implementarse con la tecnología que está en uso, cuán refinada es la especificación (es decir, cuán probable es que el cliente acepte lo que ha pedido una vez que lo vea implementado) y cuál es el esfuerzo previsto. Por eso, se suele consensuar la complejidad de cada requerimiento en *Story Points*, una unidad de medida sólo orientada a tal fin que suele expresarse siguiendo la escala de Fibonacci.<sup>16</sup>

En cambio, el esfuerzo se estima en unidades de tiempo, por ejemplo, horas/hombre. Así, el equipo tiene un modo simple de distinguir tareas complejas (cuyo esfuerzo es alto por tener muchos elementos desconocidos) de otras simples pero que demandan un trabajo muy rutinario por muchas horas. Es importante que el equipo como un todo concuerde con las estimaciones, lo cual trae varios beneficios: se enfatiza la responsabilidad grupal para cumplir las estimaciones, se minimizan las chances de subestimar las tareas y se evitan dificultades en caso de que deban

---

<sup>16</sup> La escala de Fibonacci suma los dos números anteriores para armar una secuencia progresiva. Así, sus primeros números son 1, 2, 3, 5, 8, 13, 21. Algunos métodos de estimación permiten incluir el número “infinito” para reflejar que la estimación tiene demasiada incertidumbre. El objetivo de esta escala es que el consenso del equipo técnico sea más fácil de alcanzar: en caso de divergencia sobre la complejidad de una tarea, es más probable acercar dos posturas que están debatiendo entre complejidades 8 y 13 que entre aquellos que no acuerdan entre 5 y 6.

reassignarse tareas.<sup>17</sup> Hay muchos métodos para alcanzar ese consenso; los típicos demandan involucrar a todo el equipo, y suele ser recomendable que se sumen representantes del cliente y terceros interesados.<sup>18</sup>

Una vez condensada la expectativa del cliente en *User Stories* (y sus subcomponentes), es necesario asignarle a cada *Story* una determinada Prioridad. Es decir, el equipo de desarrollo y el cliente deben tener claro cuáles de esas *User Stories* debe ser implementada en primer lugar. Hay múltiples factores a considerar, incluyendo cuestiones técnicas, otras relacionadas con el negocio y aquellas que tienen que ver con las estimaciones mencionadas en el párrafo anterior. Lo esencial es comprender que cada *Story* tendrá una determinada prioridad, que tanto el cliente como el equipo de desarrollo están de acuerdo en eso y que las que se completan primero son las que antes serán vistas, aprobadas e integradas al producto pedido por el cliente.

Llegado este punto, las Metodologías Ágiles empiezan a diferir en su dinámica. Para simplificar la exposición, se circunscribe la presentación a una de las más populares, llamada Scrum.<sup>19</sup> Siguiendo sus lineamientos, se propone consolidar al conjunto de *User Stories* ya priorizadas en una herramienta denominada *Backlog*, término que se traduce como “una lista ordenada del trabajo pendiente”. Una vez que dicho *Backlog* (o al menos una cantidad significativa de *User Stories* importantes) está completo, el equipo entonces planifica trabajar en tantas de las *Stories* de alta prioridad como entren en una primera iteración (o ciclo) de desarrollo. Este es un aspecto característico de la mayoría de las Metodologías Ágiles: la duración de cada ciclo está

---

<sup>17</sup> Para mayor detalle sobre la importancia de que las estimaciones sean consensuadas, véase Cohn (2006).

<sup>18</sup> Romeu (2014) presenta un ejemplo de Planning Poker. Otros 3 métodos son presentados por Pérez (2017).

<sup>19</sup> Un equipo de Scrum tiene autoridad, autonomía y la responsabilidad de alcanzar los objetivos de una iteración, frecuentemente bajo la tutela de un “Guía” (Scrum Master) que obra al mismo tiempo de garante de la metodología y facilitador. Se busca reflejar la propuesta de Schwaber (2004).

predefinida, y la cantidad de *Stories* se adapta a dicha restricción de tiempo. Esta es la razón por la cual se suele referir a estas Metodologías como *Time-Boxed*, a diferencia de métodos más antiguos donde el volumen de trabajo por realizar terminaba definiendo la duración del proceso operativo.

Cada iteración suele incluir muy diversas actividades: análisis de requerimientos, diseño, implementación, mantenimiento de funcionalidad existente y control de calidad – o, en otras palabras, todas las tareas necesarias para transformar una solicitud hecha en lenguaje coloquial en un producto terminado e integrado a una solución (nueva o existente).<sup>20</sup> A cada una de dichas iteraciones se las suele denominar *Sprint*, y típicamente miden entre dos y cuatro semanas de trabajo para facilitar la medición del progreso y el manejo de contingencias. Al cabo de cada iteración, el equipo de desarrollo debería haber conseguido aumentar el valor del software que produce y haber integrado nuevos elementos a un producto. A esto se refieren quienes llaman “incremental” al desarrollo bajo estas metodologías.

El avance a lo largo del *Sprint* se suele repasar diariamente en una breve reunión donde cada miembro del equipo suele describir qué objetivos cumplió el día anterior, cuál es su plan para ese día y si tiene alguna dificultad que le demande apoyo del equipo. Para alentar la brevedad y velocidad de los encuentros, la reunión se suele hacer en grupo y de pie, motivo por el cual se las conoce como *Stand-Up Meetings*. El equipo de trabajo debe tener acceso a una herramienta que condensa todas las *Tasks* que componen el *Sprint*; un ejemplo típico es un *Scrum Board*,

---

<sup>20</sup> Es importante subrayar que cada una de esas actividades debe ser realizada solamente si contribuye a implementar exitosamente cada una de las *Stories* o *Tasks*. Las Metodologías Ágiles buscan evitar lo que en los preceptos de calidad total se denomina *muda* (desperdicio): toda actividad que no contribuya directamente a entender una demanda del cliente, diseñar una solución, descomponerla en fracciones, implementarla, validarla y demostrarla.

esencialmente una pizarra (a veces física, pero generalmente virtual) donde cada *Task* y su estado (pendiente, en progreso, terminada) son fácilmente visibles.

El progreso se suele consolidar actualizando la cantidad trabajada en cada *Task* o *Story*. Si el *Scrum Board* es digital, el *Backlog* del *Sprint* se actualiza automáticamente en un gráfico cartesiano: sobre el eje Y se representa el volumen de trabajo pendiente, sobre el eje X se representan los días de trabajo que aún quedan en el *Sprint* y en el corazón del gráfico se reflejan tanto el avance planeado como el real, en forma aproximadamente lineal y de pendiente negativa. La herramienta que proporciona esta representación gráfica del avance en el trabajo es comúnmente llamada *Burn-Down Chart*.<sup>21</sup> Esta herramienta es particularmente útil para detectar desvíos, tareas obstruidas por otras que se han demorado de forma imprevista y (en términos más generales) para mitigar riesgos y amenazas de manera temprana.

Ni bien se completa un *Sprint*, el equipo de desarrollo deberá hacer una demostración del progreso al cliente, por dos razones fundamentales: probar que hubo avances y (especialmente) para darle lugar a que confirme que lo que se ha implementado está en línea con lo que efectivamente esperaba. Es importante subrayar que el cliente, bajo esta modalidad de trabajo, tiene la potestad de detallar *qué es* lo que quiere, pero debería dejarle todas las decisiones sobre *cómo hacerlo* al equipo técnico.<sup>22</sup>

Por ese motivo, durante la demostración suele ocurrir que el equipo de desarrollo y el cliente comprendan que la definición original capturada en la *User Story* era ambigua, imprecisa, parcial o incluso que ahora que ve el producto terminado el cliente ha cambiado de idea y quiere

---

<sup>21</sup> Goncalves (2019b) profundiza en recomendaciones de uso y provee ejemplos.

<sup>22</sup> Un ejemplo algo grotesco pero realista de por qué no es recomendable que tanto el “qué” como el “cómo” estén a cargo del cliente puede encontrarse en Inman (2010).

algo diferente. Puesto que el equipo habrá trabajado en las *User Stories* más prioritarias y que habrá hecho sólo lo necesario para satisfacerlas, ninguna de esas instancias debe implicar un problema. De hecho, el propósito principal de las Metodologías Ágiles es permitir flexibilidad ante los cambios: si cualquier error o desvío en las *User Stories* más prioritarias para el cliente es detectado lo antes posible, el impacto en el costo total del proyecto será más acotado.<sup>23</sup> Como ventaja adicional, el equipo reaccionará al instante si las necesidades del negocio del cliente se modificaran durante el proceso.

Como buena práctica, antes de pasar a la siguiente iteración de desarrollo, los equipos (con o sin presencia de representantes del cliente, aunque lo último es más recomendable) suelen tener una reunión conocida como *retrospectiva*, donde analizan qué comportamientos deben ser mejorados, cuáles deben mantenerse y cuáles deben abandonarse para que el siguiente ciclo sea más fructífero que el anterior. De tal modo, no solamente el software incrementa su valor; idealmente, los problemas de procesos serán detectados y corregidos rápidamente y la sinergia del equipo será mayor tras cada ciclo que se completa. Como la responsabilidad de cada entregable al final de cada ciclo es de todos y cada uno de los miembros del equipo, esta instancia de autoanálisis es esencial para que el grupo de desarrollo tenga éxito a largo plazo.

Como se mencionó anteriormente, algunos elementos de la dinámica aquí descrita son exclusivos de la metodología *Scrum*. No obstante, prácticamente todas las Metodologías Ágiles están planteadas para funcionar a través de ciclos que comprenden planificación, análisis, diseño, programación, pruebas, finalización, demostración y documentación. Si bien cada ciclo tiene sus

---

<sup>23</sup> Tanto la entrega frecuente (para discutir cara a cara con el cliente su grado de satisfacción) como la recomendación de implementar con simpleza (es decir, de hacer solamente lo requerido sin desperdiciar esfuerzo en otras tareas) son esenciales para entender a las metodologías Ágiles. Véase Beck, Beedle, van Bennekum *et al* (2001), especialmente los Principios 1, 2 y 10.

particularidades (ya sea por el desafío técnico puntual que se intenta resolver, por la curva de aprendizaje de una nueva tecnología o por los imponderables que puedan presentársele a los miembros del equipo), conforme crece la cantidad de ciclos pasados se vuelve viable recopilar y procesar información capturada y a través de ello refinar estimaciones sobre qué puede ocurrir en el futuro.

La información puede ser traducida a métricas que se convierten, a su vez, en herramientas descriptivas y reflexivas que asisten a los equipos en la planificación de futuros ciclos de desarrollo. Se detallan algunas de ellas en el siguiente apartado.

## II.B. Métricas

Presentada ya la dinámica general de las Metodologías Ágiles de Desarrollo de Software, se vuelve preciso plantear una cuestión trascendental: es recomendable tener un modo de medir el rendimiento del equipo de desarrollo. El éxito del equipo puede evaluarse en dos dimensiones luego de cada cierre de *Sprint*: la del éxito del proyecto (evaluando tiempo, costos, calidad de las especificaciones, satisfacción de terceros interesados, métricas de desarrollo y métricas de calidad) y la del éxito del producto (lecciones aprendidas sobre la dinámica de trabajo, la implementación exitosa de tecnología innovadora, el logro de objetivos de la organización y la posibilidad de que el producto se siga usando en el futuro).<sup>24</sup>

Las Métricas Grupales más importantes para medir los procesos incluyen:

- Esfuerzo: la sumatoria de horas invertidas en actividades del proyecto durante el último ciclo.

---

<sup>24</sup> Podjavo y Berzisa (2017), pps. 1 y 2.

- Productividad: es el resultado de dividir una unidad de tamaño (Líneas de Código, *Story Points*, o cualquier otra) por la cantidad de horas expresadas por la métrica Esfuerzo.
- Duración del Proyecto: es la duración total de un proyecto en días sin contar aquellas jornadas libres, festivas o inactivas.
- Predictibilidad: mide la variación entre la duración estimada del proyecto y la real.
- Tasa de Requerimientos Completos: para un período de trabajo determinado (por ejemplo, para cada Sprint), representa qué porcentaje de los requerimientos planificados pudieron efectivamente completar el ciclo de desarrollo.
- Densidad de Defectos en Producción: representa la relación entre la cantidad de defectos únicos encontrados en el sitio productivo durante un período de tiempo (por ejemplo, 6 meses) y una unidad de tamaño (*Story Points*).<sup>25</sup>
- Uptime/Downtime: refiere a la cantidad de horas en que un producto ha estado disponible para que lo utilicen usuarios finales.<sup>26</sup> Si el equipo ha producido código con limitaciones de *performance* o la tasa de defectos no detectados es muy alta, es seguro que tendrán que ejercer mantenimiento periódico al producto (fuera de los plazos planificados originalmente) para resolver esos problemas. Idealmente, *Uptime* (el tiempo en funcionamiento normal) tiene que tender al tiempo calendario transcurrido en el proceso bajo análisis y *Downtime* (tiempo fuera de servicio) debe tender a cero.

---

<sup>25</sup> Un Defecto es, esencialmente, un desvío en el comportamiento de una aplicación que impide cumplir con un requerimiento.

<sup>26</sup> Es especialmente fácil entender el concepto cuando se piensa en una página web y el tiempo en que permanece accesible.

- Velocidad del Equipo: quizás la más utilizada para estimaciones, representa cuántos *Story Points* planea resolver un equipo de desarrollo por Sprint promedio. Suele basarse en información histórica siempre que está disponible.

A nivel individual muchas de estas métricas pueden replicarse. El objetivo es interpretar desvíos entre rendimientos individuales o ponderar la importancia de las tareas realizadas. Es importante, sin embargo, subrayar que las métricas individuales jamás pueden interpretarse sin interiorizarse en los factores cualitativos que se desprenden del trabajo grupal.

- Productividad por Sprint: se trata de la relación entre la cantidad de *Story Points* que un determinado programador ha logrado completar durante un *Sprint* y las planificadas para todo el equipo originalmente.
- Varianza Productiva: la diferencia entre la productividad esperada al momento de planear el *Sprint* (calculada como el total del *Backlog* asignado a cada programador dividido por la velocidad promedio del equipo) y la que alcanzó efectivamente el programador durante un determinado *Sprint*.
- Índice de Productividad: similar al anterior, será el cociente entre la cantidad de tareas efectivamente completadas por un desarrollador durante un *Sprint* y las planificadas originalmente. Si dicho cociente fuera cercano o superior a 1, indicará una alta performance o una subestimación de la complejidad original.

Todas estas métricas individuales se desprenden del concepto de desarrollo bajo Metodologías Ágiles. No obstante, desde una perspectiva técnica, hay métricas mucho más sencillas de obtener: aquellas que se producen a través del trabajo de cualquier programador bajo cualquier metodología. Los programadores que trabajan en un proyecto colectivo esencialmente

interpretan requerimientos, escriben líneas de código para resolver los problemas que ellos plantean, promueven dichas líneas de código a un repositorio compartido con otros programadores y las modifican posteriormente si fuera necesario. La eficacia de cada programador (o de ellos en conjunto) también puede caracterizarse, entonces, interpretando la siguiente información:

- Cantidad de Líneas de Código: como su nombre lo indica, condensa la cantidad de líneas de código producidas por cada programador. Se trata de una medida extremadamente rústica, en tanto no expresa por sí sola si el diseño es bueno o malo o la complejidad de la tarea relacionada con el código producido. No obstante, un número extremo (por alto o ínfimo) suele ser un buen llamado de atención.
- Commits/Check-Ins: esta operación, de difícil traducción, refiere a la cantidad de veces que un programador promueve una unidad atómica de código al repositorio común. Es una medida más comprensiva del avance de un programador, ya independientemente de la cantidad de líneas que haya producido.
- Pull Requests: es el modo en que un programador propone incorporar un cambio molecular (es decir, compuesto por varios átomos o *check-ins*) a un repositorio principal. Con frecuencia, una *Pull Request* (una vez aprobada) marca la compleción de una tarea, la resolución de un defecto o la finalización del desarrollo de una *Story*.
- Eficiencia del Código: la proporción (generalmente expresada como porcentaje) de código producido por un programador que se mantiene cuando se integra una solución al producto terminado. Típicamente, si el análisis y diseño son correctos y la implementación es exitosa, habrá muy poco código (y tiempo) desperdiciado. Esta medida es difícil de obtener y destilar (especialmente cuando ha habido cambios en los requerimientos del cliente) dado que demanda análisis técnico del código en el repositorio.

- Tasa de Código Improductivo: es el reverso del indicador anterior. Es decir, el porcentaje de código producido y subido a un repositorio que es abandonado o modificado antes del fin de la implementación.
- Flexibilidad del Código: vinculada a las anteriores, esta métrica busca estimar la cantidad de retrabajo en horas que demanda cada nueva *feature* a medida que se agregan nuevos requerimientos. Esta métrica se vuelve un problema frecuente en productos cuyo código productivo tiene varios años de construido. Por ese motivo, esta métrica requiere de análisis técnico para ser producida e imputada correctamente.
- Deuda Técnica: puesto resumidamente, es un concepto que refleja el costo de retrabajo futuro que genera una solución rápida para un problema urgente de hoy.<sup>27</sup> Refleja la tensión entre lo que se puede alcanzar de modo realista en un tiempo limitado. En teoría, se la puede erradicar con altas dosis de conocimiento, capacidad técnica, buenas prácticas de diseño y la capacidad para persuadir al cliente sobre la importancia de realizar tareas de mantenimiento sin beneficios visibles para prevenir problemas y costos futuros. En la práctica, no existe proyecto duradero que no la padezca en mayor o menor grado y es especialmente difícil distinguir entre la Deuda Técnica que obedecía a la complejidad inherente del proyecto (o su contexto) y la que fue agregada consciente o involuntariamente por el equipo de desarrollo.
- Peer-Code Reviews: esta métrica cualitativa se produce a través del análisis de pares del grupo de trabajo. La inspección del código común suele generar propuestas, críticas y una

---

<sup>27</sup> Esta definición no abarca todos los escenarios posibles, para simplificar la comprensión. Un ejemplo adicional: también puede surgir deuda técnica de fenómenos ajenos al equipo de desarrollo, especialmente cuando un producto se tiene que integrar con soluciones provistas por otros proveedores. Por caso, cuando Microsoft publica un nuevo navegador web (como Microsoft Edge) puede que una página empiece a mostrar defectos cosméticos cuando un cliente trata de acceder a ella usando ese nuevo navegador. La necesidad de reparar esos problemas puede no ser inmediata, pero deberá considerarse deuda técnica e integrarse a la lista de tareas pendientes.

lista de modificaciones sugeridas. Si la naturaleza de las observaciones sobre el código de un mismo programador tiende a repetirse en el tiempo, es posible que esté señalando un área de mejora o la necesidad de entrenamiento técnico.

No obstante, la medición de la productividad individual no puede dissociarse de las métricas de calidad para no correr riesgos de confundir velocidad con rendimiento destacado. Las siguientes métricas contribuyen a distinguir entre ambos conceptos:

- Densidad o Cantidad de Defectos en Producción: esta métrica (anteriormente expresada a nivel grupal) puede atribuirse a programadores individuales cuando se identifica que parte del código que han producido fue responsable del fallo detectado en el producto final.<sup>28</sup>
- Vida Media de los Defectos: expresa el tiempo promedio que demora en repararse un defecto reportado a lo largo de todos los *Sprints*. Un número elevado puede expresar múltiples síntomas: puede que el encargado de reparar el defecto esté atendiendo otras tareas prioritarias, que no esté atendiendo debidamente los defectos o que suela incurrir en errores de diseño que (una vez señalados por el defecto) le resulte muy costoso resolver con una nueva implementación.
- Tiempo de Ciclo de las *Stories*: expresa el tiempo promedio que lleva que la idea condensada en una *Story* sea promovida a Producto Final, incluyendo el tiempo que toma probarla y validarla.

---

<sup>28</sup> Es importante subrayar que esta métrica también puede penalizar a las prácticas de control de calidad que no hayan detectado el problema pese a haber validado el funcionamiento de esa *Story*. Siguiendo preceptos de calidad total, se trata menos de encontrar un culpable de la situación que de identificar la causa raíz para remediar el proceso y evitar otra ocurrencia. La responsabilidad es siempre colectiva.

- Tasa de Demora: la cantidad de *Story Points* que en promedio necesita ser pospuesta para *Sprints* posteriores producto de errores en las estimaciones o demoras en la implementación de otras *Stories* más prioritarias.
- Cantidad de Defectos Reportados por cada Story/Story Point: como su nombre lo indica, busca (a nivel individual o grupal) la relación entre la cantidad de defectos reportados por cada *Story* o cantidad promedio de *Story Points*. Generalmente, es una métrica útil cuando presenta variaciones bruscas respecto de promedios históricos.

Por último, la dinámica del equipo suele incidir directamente en el éxito del proyecto dado que el grupo de trabajo es el principal agente en la implementación. Las mediciones tienen un componente esencialmente cualitativo<sup>29</sup> que se suele evaluar individualmente operacionalizando las siguientes categorías como Comportamientos Individuales:

- Habilidades Sociales: las Metodologías Ágiles dependen de múltiples factores colaborativos para dar el mejor resultado. La adherencia a los procesos (incluyendo puntualidad para asistir a las reuniones pautadas y la eficacia de la participación), la colaboración con miembros del equipo (por ejemplo, proveyendo una solución técnica específica o revisando el producto del trabajo ajeno) y la comprensión de los procesos para proponer mejoras son algunos elementos que se suelen evaluar.
- Colaboración: los desarrolladores deben idealmente contribuir unos con otros, incluyendo tareas de *coaching*, entrenamiento, compartir ideas, contribuir con soluciones creativas y transmitir conocimiento técnico o analítico.

---

<sup>29</sup> Berzisa y Podjavo (2017), en cambio, proponen cuantificar métricas sociales que reflejen la cohesión del equipo, la naturaleza de la interacción y la frecuencia con que cada miembro del equipo valora positiva o negativamente a su contraparte.

- Amplitud de Conocimiento: las tareas en grupos de Desarrollo Ágil se distribuyen entre todos los miembros. Por ese motivo, los participantes ultra especializados en pocas áreas contribuyen menos al éxito del equipo que los que son sólidos en una diversidad de tareas (programación, control de calidad, diseño, análisis y especialista en visión del cliente). Esta métrica busca recompensar la adaptabilidad, la proactividad y la capacidad de desarrollar múltiples tareas con destreza comparable.

## II.C. Métricas y Evaluaciones: cómo evitar que devoren al equipo

En los apartados anteriores se ha descrito la dinámica colectiva de producción de Software siguiendo Metodologías Ágiles, las definiciones de éxito para cada equipo de trabajo y el significativo volumen de información de métricas tanto grupales como individuales que puede recopilarse.

Como puede apreciarse, hay una cantidad enorme de información que se desprende de las Metodologías Ágiles de Desarrollo de Software, la dinámica de grupo y la propia programación de Software (bajo cualquier metodología). Como ventaja adicional, nuevas herramientas y soluciones tecnológicas surgen continuamente, lo cual facilita la recolección e interpretación de datos. Desde ya, la interpretación de las métricas aisladas (individuales o grupales) no es recomendable, en tanto no se las puede comprender sin contexto. Buscar remediarlas sin contar con un nivel de análisis puede generar efectos contrarios a los deseados, dado que el equipo técnico tendrá elementos para *engañar al sistema*.

Por otra parte, conforme el tamaño de los equipos de desarrollo aumenta, el desafío de interpretar los datos (en especial a nivel individual) demanda mucho más tiempo y participación de los recursos que cuentan con elevada jerarquía técnica, tienen mayor familiaridad con el

producto y aptitud para dar *feedback* constructivo a sus pares. Estas características no siempre son sencillas de encontrar (o de fomentar) en los programadores.

Esto último implica un problema adicional: procesar la información se vuelve esencial en contextos donde hay casos límite o bien donde el grupo de desarrollo ha rendido por debajo de las expectativas. Pero es justamente en esos escenarios donde los recursos técnicos mejor capacitados estarán abocados a recomponer las expectativas del cliente, quizás sumándose al equipo de trabajo para resolver problemas puramente técnicos. En la práctica, por lo tanto, la información más importante suele ser difícil de obtener en los momentos más álgidos, y se requiere de disciplina en el liderazgo y el Management para que el equipo haga los diagnósticos adecuados una vez superadas las crisis.

Hay fundamentos, incluso, para recelar de la pertinencia de recolectar cualquiera de las métricas individuales y grupales que la propia metodología provee. Por un lado, porque la bienintencionada definición de qué métricas se capturan suele alterar el comportamiento de los evaluados, lo que se conoce como la paradoja del observador. Por otro, porque la capacidad de elegir las métricas que mejor resumen el rendimiento no es un talento para nada trivial. Muchas veces el responsable de diagnosticar las causas de un problema tiende a elegir métricas irrelevantes o que agravan la dificultad, solamente porque son las que más conoce o le resultan más accesibles.

En este contexto, McHugh (2012) considera que la única recomendación razonable parece ser limitarse a relevar la menor cantidad de métricas posible y sólo dar crédito a las referidas al éxito del cliente/usuario final. Esa única medida de la performance del equipo puede circunscribirse, por ejemplo, a relevar el cumplimiento de fechas y presupuesto y valor de negocio

realizado, siendo esto último entendido como cantidad de *Story Points* entregados desde la perspectiva del negocio (en vez de la del equipo).

La ventaja de ese enfoque es que le da elementos al propio equipo para remediar los problemas de procesos que impactan negativamente en la satisfacción del cliente. Es concebible que uno o varios miembros de un grupo estén trabajando por debajo del nivel óptimo: puede que no estén utilizando el máximo de su capacidad, que no comprendan el negocio, que no se adapten bien a la dinámica del equipo o que no hayan conseguido incorporar la visión estratégica del proyecto. En cualquiera de esos casos, dado que la responsabilidad de satisfacer al cliente es del equipo en su conjunto, el propio equipo debe hablar del problema, elegir (e incluso crear) las métricas que permitirán medir si la brecha entre el rendimiento esperado y el alcanzado se está cerrando y hacer un seguimiento del problema. Provenzano (2018) sugiere que si el precio de un cliente satisfecho es sacrificar toda posibilidad de evaluar el rendimiento individual (a menos que el propio grupo solicite reemplazar a uno de sus miembros), el abandono de métricas es preferible.

Es posible también adoptar posturas intermedias. Lowe (2019) postula que, si se parte de la base de que cada miembro del equipo, cada componente, cada grupo de trabajo y cada proyecto es *como un copo de nieve* (es decir, únicos en su complejidad más allá de su apariencia similar), eso quiere decir que las métricas nunca pueden ser medios válidos de comparación. Para peor, el hecho de que cuando se siguen Metodologías Ágiles el aprendizaje es una de las ganancias más valiosas, no siempre es adecuado mensurar el éxito según la adherencia a un plan o a un proyecto; en esencia, la satisfacción del cliente puede haberse alcanzado incluso habiendo fallado en esos aspectos.

Entonces, ¿qué nos aportan las métricas en esta perspectiva? Esencialmente, nos ayudan a construir hipótesis de trabajo y a alcanzar la satisfacción del cliente lo antes posible. Por ejemplo, si una métrica nos indica que un alto porcentaje de clientes abandona una compra al llegar al carrito, puede deberse a que la reacción del carrito es demasiado lenta, a que el diseño es poco intuitivo o sencillamente que las condiciones de pago son desventajosas respecto a un competidor. Las dos primeras hipótesis deben ser resueltas por el equipo de desarrollo, pero la tercera entra en la órbita del negocio.

Lo que está al alcance del equipo de desarrollo es descartar los problemas que se consideran más probables hasta que efectivamente el producto consiga revertir el comportamiento de sus usuarios. Algo similar ocurre cuando se evidencia la necesidad de revisar los procesos de trabajo del equipo: una cantidad inusual de defectos no detectados puede tener múltiples posibles causas raíz, y lo mejor que las métricas pueden hacer es señalar que hay un interrogante que desenmarañar.

Una inquietud fundamental se presenta en este punto: si es preciso ser tan cuidadoso con la interpretación de la información que el propio equipo de trabajo genera diariamente, ¿qué sentido tienen las evaluaciones de desempeño individuales impulsadas por la gerencia de RRHH? Dicho de otro modo: ¿qué herramientas superadoras a la propia dinámica de las Metodologías Ágiles tiene la empresa para alinear la performance de sus equipos de desarrollo, dado que cada uno de ellos está siendo constantemente evaluado por los clientes y usuarios finales y que la productividad de cada individuo tiene máxima visibilidad en el seno de su equipo de trabajo?

La cuestión es atendible, en tanto hay motivos de sobra para ser escépticos respecto a la posibilidad de que cualquier método de evaluación de un equipo (que exceda la dinámica de las reuniones retrospectivas presentadas anteriormente) aporte algo más que lo que daña.

La necesidad de hacer distinciones entre rendimientos individuales (especialmente si esas decisiones implicarán nuevas oportunidades de desarrollo o un cambio en la compensación de uno, varios o todos miembros) afecta negativamente la moral del equipo y conlleva una cantidad significativa de disfunciones futuras.<sup>30</sup> Para empezar, la mera sospecha de que existe un sistema de reconocimiento individual generará prácticas competitivas, que erosionan la colaboración. Luego, puede verse afectada la relación entre ese grupo y otros clientes internos en caso de que el reconocimiento sea universal para el equipo, pero no se extienda a sus pares.

Además, como estrategia motivacional, un reconocimiento (incluso si fuera un recurso para empujar al equipo a conseguir un objetivo demasiado ambicioso) puede ser percibido como un gesto de cinismo, generando el efecto contrario y dañando la identificación de los empleados con la empresa. Incluso si el reconocimiento individual buscara incentivar el desarrollo de un área específica, es perfectamente posible que el proceso como un todo se resienta.<sup>31</sup> Finalmente, existe el riesgo de que una vez que la dinámica de trabajo pone como norte una calificación individual o un reconocimiento económico, puede que se destruya la motivación de hacer un trabajo para satisfacer los propios valores y objetivos.

Todos estos potenciales efectos negativos no han implicado que las empresas especializadas en el Desarrollo de Software hayan abandonado los Sistemas de Gestión de Desempeño. Antes bien, es probable que muchos de esos sistemas se encuentren en transición, con la esperanza de redefinirlos para que permitan alinear los objetivos de los diferentes equipos de desarrollo con la estrategia de la empresa.

---

<sup>30</sup> Poppendieck (2004) presenta en detalle un caso práctico en su trabajo *Unjust Desserts?*

<sup>31</sup> Ejemplos típicos son los premios por defecto encontrado (que terminan por dañar la cooperación entre los desarrolladores y los *testers*) y los premios por alcanzar una fecha demasiado ambiciosa (que suelen derivar en una peor calidad del código entregado, mayor esfuerzo de soporte y alta insatisfacción en el usuario final).

A lo largo del Cuerpo Teórico se caracterizó a las Herramientas de Evaluación de Desempeño y a las Metodologías Ágiles de Desarrollo de Software. Las primeras (especialmente cuando su implementación es exitosa) son susceptibles de capturar la contribución individual de cada miembro de la empresa. Entre aquellas menos estructuradas y de creciente popularidad se cuentan Objetivos de Desempeño y Observación de Comportamientos. No obstante, ellas no están exentas de tensiones y dificultades, especialmente aquellas que se desprenden de la subjetividad del evaluador.

Se postula que las Metodologías Ágiles de Desarrollo de Software pueden enriquecer y potenciar el funcionamiento de las Herramientas de Gestión de Desempeño; esto puede darse a través de la documentación relacionada con el refinamiento de los procesos o áreas individuales de mejora, y también gracias a la consolidación de métricas y observables que permitan aumentar la objetividad del proceso.

En el Cuerpo Práctico del presente trabajo se indaga sobre el grado y el modo en que se vinculan ambas formas de gestión en empresas de desarrollo de software con sede en el Área Metropolitana de la Ciudad de Buenos Aires.

## Cuerpo Empírico

El Cuerpo Empírico del presente trabajo procura caracterizar a las Herramientas de Evaluación de Desempeño Individuales que se utilizan hoy día en equipos de desarrollo de Software Ágil, como así también la forma de implementar esas Metodologías y qué métricas grupales e individuales se capturan cotidianamente. La información recolectada permite explorar la relación entre cada sistema y determinar si la integración exitosa entre ambos cuerpos de información es un factor que suele ir de la mano con una valoración positiva respecto a ambas implementaciones.

A tal fin, a lo largo del Capítulo III se relevan 14 empresas del sector de Desarrollo de Software del Área Metropolitana de Buenos Aires. Primeramente, se indaga sobre la implementación de Herramientas de Evaluación de Desempeño que tienen actualmente en funciones: la frecuencia con la que se celebran reuniones de *feedback* y las métricas u observables que se discuten; los métodos de evaluación que cada empresa usa y el modo en que se los calibra; por último, se procura delinear el grado de involucramiento de la dirección en el proceso.

Luego, en lo que respecta a Metodologías Ágiles, se inquiriere sobre las peculiaridades de cada implementación, las métricas que se recolectan (sean relacionadas con la Metodología y el Negocio, con Web/Application Development o con la Calidad del Código) y con qué nivel de detalle o agregación (individual o grupal).

Posteriormente, se busca averiguar en qué instancia esas dos masas de información se integran o consolidan, cuáles son los fines prácticos percibidos en dicho proceso, en qué grado se los alcanza y qué grado de satisfacción declara cada entrevistado para con el sistema actual. Acto

seguido, se toma nota de cuál sistema entre los abordados tiene mayor relevancia a la hora de tomar decisiones operativas, sean promociones, decisiones en torno al salario o planes de carrera.

Finalmente, en el Capítulo IV se desarrolla una serie de recomendaciones para mejorar la articulación de la información y la gestión de equipos de desarrollo que tienen Herramientas de Evaluación del Desempeño en funciones. Dichas recomendaciones están basadas en las prácticas seguidas por aquellos respondientes que se manifiestan más conformes con sus implementaciones actuales, como así también en los defectos y limitaciones identificados por aquellos respondientes que han identificado problemas y áreas de mejora.

## **Capítulo III: La relación práctica entre las Herramientas y los Equipos**

### **III.A. Metodología y Participantes de la Muestra**

El propósito de este apartado es descriptivo y el nivel de análisis será cualitativo.

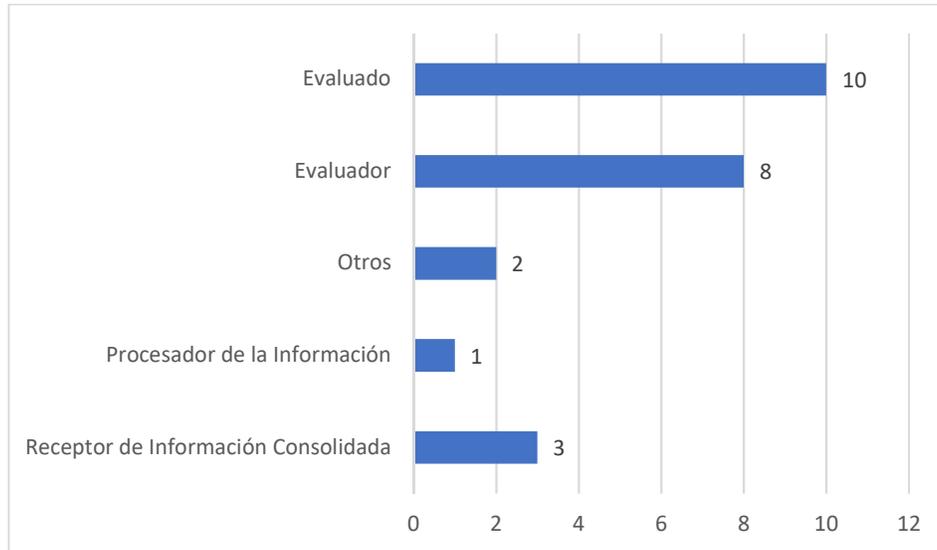
Los contenidos de este capítulo presentan la información recabada durante entrevistas guiadas hechas a representantes o consultores que trabajan en 14 empresas con al menos una sede en la Ciudad de Buenos Aires y que cuentan con equipos de desarrollo propio. Algunas de dichas empresas tienen equipos de desarrollo híbridos que poseen exclusivamente representación local, mientras otros interactúan con otras sedes que la misma empresa tiene en distintas locaciones. Más allá de las posibles diferencias culturales, se ha observado en los casos relevados que la dinámica del sistema de gestión de desempeño y las herramientas utilizadas para evaluarlo son las mismas en todas las locaciones.

Las entrevistas guiadas recopilan información exhaustiva sobre cada una de las implementaciones, en línea con el objetivo descriptivo del presente capítulo. No obstante, la interpretación de la información obtenida incluye en ocasiones la distribución de las respuestas y las características de las empresas representadas por cada respondiente. La pretensión no es, sin embargo, generalizar las conclusiones al universo de empresas del rubro ni producir una muestra estadística significativa.

### **III.B. Caracterización de los Respondientes (por Rol y Tamaño de Empresa)**

5 de los entrevistados son solamente evaluados con la herramienta sobre la que han respondido, mientras que 2 son solamente evaluadores. Los 7 respondientes restantes se desempeñan como evaluadores, evaluados, facilitadores del proceso con un rol de consultores

externos o bien consolidadores y receptores de la información recopilada por otros miembros de su organización.



*Figura 1 – Caracterización de los Respondientes (por Rol)*

Algunos de los evaluadores tienen responsabilidades formales en gestión del talento, motivo por el cual su seguimiento periódico y frecuente a los evaluados es parte de su responsabilidad (y a su vez un criterio por los que ellos mismos eran evaluados). Especialmente en esos casos, en los resultados se separa el proceso formal de evaluación del desempeño de las instancias intermedias de seguimiento.

Puesto que el objetivo de la presente investigación se centra en el vínculo entre la dinámica cotidiana de desarrollo de software y las Herramientas de Evaluación de Desempeño, se considera que la información más valiosa la pueden proveer empresas que tienen una Herramienta de Evaluación del Desempeño formal. Pese a ello, es una sorpresa descubrir que 13 de las 14 empresas relevadas (92.86%) tienen una en funcionamiento, mientras que la restante empresa (7.14%) la

tuvo en el pasado y la reemplazó por un proceso infrecuente y no estructurado.<sup>32</sup> Se predecía una distribución menos uniforme, teniendo en cuenta que (como se discutió en el Capítulo I) gestionar dichas herramientas es costoso y que su aporte suele ser materia de discusión.

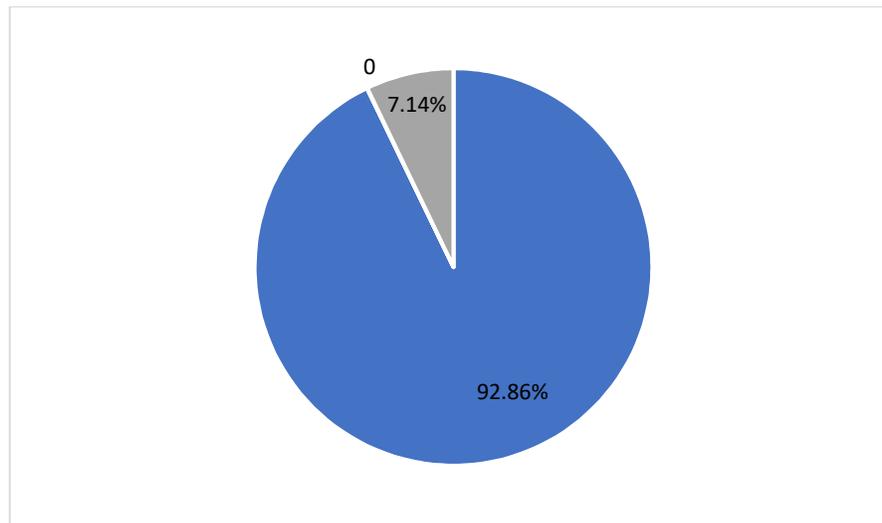


Figura 2 - Existencia de una Herramienta de Evaluación del Desempeño en Funciones

Respecto de los tamaños de las empresas relevadas y de sus equipos de desarrollo, se han incluido en el análisis compañías de más de cien equipos de desarrollo, otras que solamente tienen uno y un tercer grupo que se sitúa a medio camino. Como podía intuirse, los equipos de desarrollo están mayoritariamente compuestos por programadores; en algunos casos, equipos suficientemente grandes son conformados además por Gerentes de Proyecto, Analistas de Control de Calidad, Automatizadores de Pruebas, Analistas Funcionales, *Scrum Masters* y/o Diseñadores. Lógicamente, en estos casos la dinámica cotidiana de trabajo demanda de ellos una mayor interacción con clientes y proveedores internos, pero se destaca que incluso así los programadores

---

<sup>32</sup> Parte de los motivos que explican el abandono de la previa herramienta de gestión de desempeño fue la reducción de la planta de empleados por el hecho de que la empresa está atravesando dificultades financieras. La renuncia de buena parte de la gerencia y del anterior equipo de liderazgo ha achatado la estructura de la empresa. Por otra parte, las limitaciones económicas han espaciado los ajustes salariales. La combinación de todos estos factores parece explicar por qué los equipos operativos han tomado con naturalidad (o resignación) los cambios en la dinámica.

son mayoría. En 11 de los 14 de los casos relevados (78.57%), los programadores son al menos dos tercios del total de miembros del equipo.

### III.C. Herramientas de Evaluación del Desempeño

#### III.C.I. Frecuencia de Reuniones de Feedback y Métricas Capturadas

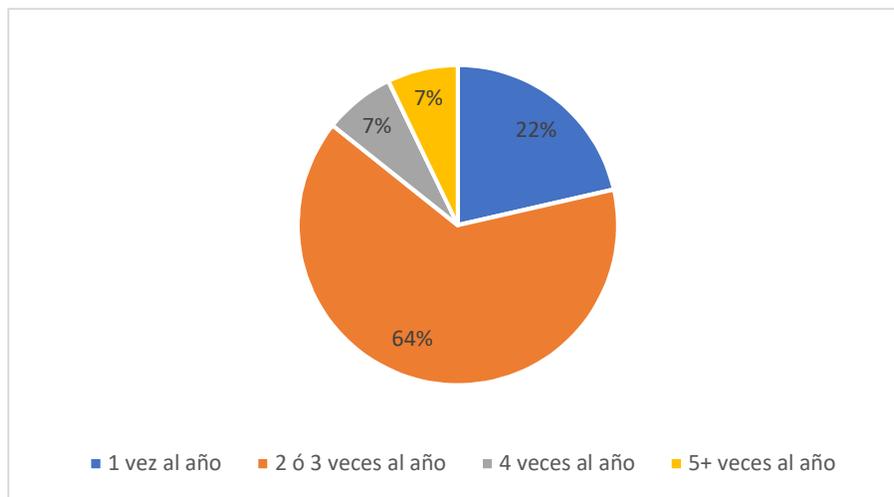


Figura 3 - Frecuencia de las Reuniones de Feedback

Como se mencionó anteriormente, las **14** empresas consultadas tienen (o habían tenido hasta meses atrás) una Herramienta de Evaluación del Desempeño. La frecuencia con la que formalmente se utiliza dicha herramienta en cada una de las empresas se revela como bastante uniforme: en **9** de los **14** casos relevados (64%) la instancia formal de evaluación (especialmente a través de una reunión de *feedback*) se da con frecuencia semestral o cuatrimestral, para un total de entre 2 y 3 ocasiones al año. De las restantes, 3 empresas (22%) optan por hacer una sola reunión formal al año, 1 (7%) lo hace con periodicidad trimestral y la restante opta por hacer evaluaciones formales con frecuencia bimestral.

En los casos situados en los extremos de la escala de frecuencias, la elección de la periodicidad con la que se realizan instancias formales de evaluación se ve correctamente alineadas

con los objetivos de cada herramienta. En otras palabras, las evaluaciones que buscan principalmente delinear planes de desarrollo y medir la adquisición de nuevas habilidades son las que celebran instancias formales de evaluación con mayor frecuencia. En la posición opuesta, las empresas que optan por realizarlas anualmente tienen como principal propósito fijar objetivos individuales o arribar a una calificación general, derivándose de esta última los ajustes salariales. Por último, en las **9** empresas cuyas instancias de evaluación se dan cada seis o cuatro meses los propósitos (y medios para alcanzarlos) son más heterogéneos, lo cual es esperable dada la relativamente alta frecuencia de esa respuesta.

En cuanto a las métricas relevadas, todas las cuales han sido mencionadas en el Capítulo I del presente trabajo, en **10** de las empresas se consolida una Calificación (general o total) durante cada evaluación, en **7** de ellas se definen Objetivos Individuales y en **11** de ellas se busca definir o evaluar Competencias y Comportamientos Individuales. Con frecuencia menor se observan empresas que documentan objetivos que se desprenden formalmente o cascadean informalmente de la estrategia de la empresa (**6/2** respondientes, respectivamente) y la ponderación de objetivos o competencias por tipo (**4** respondientes). Las empresas relevadas también buscan medir la capacidad de liderazgo de empleados específicos, ya sea con impacto en su remuneración (**2** casos) o sin hacerlo (**6** casos). Estas frecuencias son coherentes con la adopción de herramientas de evaluación que focalizan en gestionar comportamientos y objetivos, progresivamente adoptados por empresas que dependen de ser exitosas durante la detección del talento y preservación de su capital humano.

Entre las métricas y herramientas menos adoptadas, solamente 1 de los respondientes declara que se ponderan objetivos y comportamientos uno a uno. Ninguna de las empresas relevadas confecciona o consolida la información en una *Balanced Scorecard*; esta última es una

herramienta que según los respondientes no aporta mucho valor extra y requiere mayor esfuerzo en el procesamiento de la información.

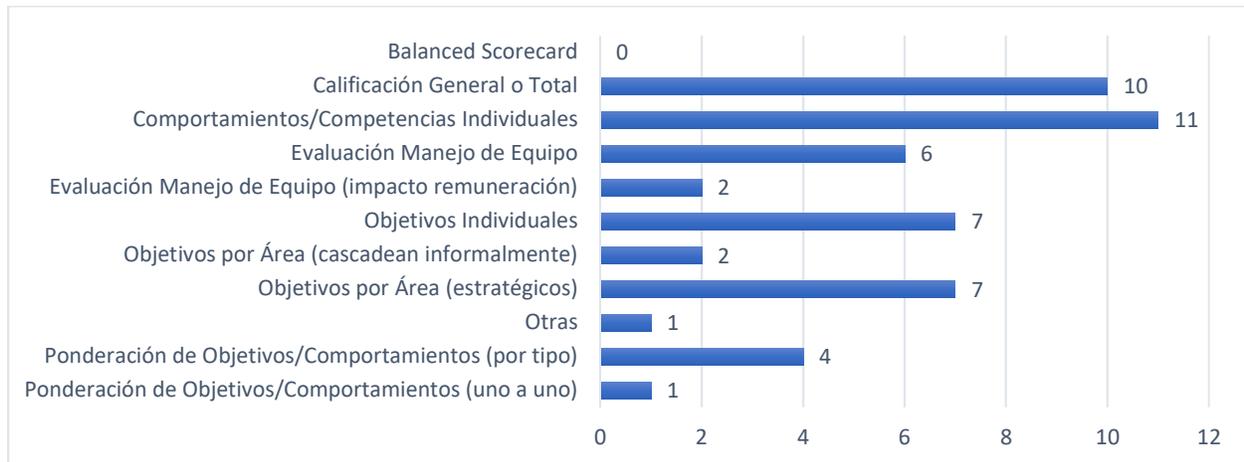


Figura 4 - Métricas/Herramientas Relevadas Individualmente

Finalmente, los respondientes subrayan que hay diversas métricas que se documentan en la evaluación de desempeño por fuera del formulario: por ejemplo, en forma de comentarios o una sección independiente de los comportamientos y objetivos para toda la empresa (e incluso otros equipos de desarrollo). Dichas métricas suelen tener relación directa con observables, que a su vez se vinculan con la rentabilidad de la empresa, la rentabilidad del cliente, la sustentabilidad de las relaciones laborales o una combinación de todos estos aspectos. Entre estos valores ad-hoc se dan ejemplos como *feedback* técnico del cliente, la evaluación periódica del clima laboral, objetivos puntuales para asegurar el acatamiento de normas ISO, el grado de cumplimiento de *Service Level Agreements* acordados con cada cliente y *actionables* (o KPIs) específicos solamente para algunos equipos dedicados a proyectos o clientes particulares.

Así, estas métricas procuran subsanar los huecos y faltantes que se suelen presentar en los formularios de herramientas que deben ser utilizadas por equipos de desarrollo muy diversos (e incluso por toda la compañía), en especial cuando es preciso cuantificar responsabilidades y

objetivos que reflejan la percepción que tienen los clientes externos del equipo bajo análisis. Algunas herramientas deliberadamente dejan espacios en blanco que son utilizados en estas circunstancias: 2 respondientes expresan que hay en sus evaluaciones espacio para objetivos comunes a la empresa (predefinidos), objetivos grupales (predefinidos), objetivos individuales (libres) y evaluación por comportamientos (con lineamientos laxos sobre los que evaluador y evaluado pueden intervenir). En estas dos ocasiones se da, entonces, una combinación de todos los elementos mencionados en este apartado.

### III.C.II. Métodos y Calibración

Respecto de los métodos utilizados en los casos relevados, resulta notorio el avance de aquellos que ponen foco en el empleado y procuran garantizar la consolidación de *feedback* de diversas fuentes. Los más utilizados son las Autoevaluaciones (**10** casos) y Evaluaciones 360 informales (**7** y **2** casos respectivamente; en **1** de estos últimos casos el proceso aún es formal, pero está en consideración flexibilizarlo). Es preciso aclarar que existen empresas donde la Evaluación 360 informal incluye documentación de la autoevaluación, y otras donde ese paso es reemplazado por una instancia más informal durante la instancia de *feedback* con el evaluador.

Con menor adopción que estos métodos se incluye la calibración de calificaciones según criterios predefinidos (**5** casos). No parece sorprendente que ese método continúe siendo relevante sobre todo en empresas que emplean a una plantilla numerosa, siendo que es precisamente en ellas donde la cantidad de evaluados y evaluadores trae una mayor necesidad de ejecutar una calibración de calificaciones. La estrategia busca uniformizar criterios de los evaluadores a la vez que incidir en la distribución de las calificaciones dentro de un grupo, de modo tal que no todos los evaluados puedan recibir calificaciones idénticas. Más adelante se afirma que esa calibración suele no ser estricta.

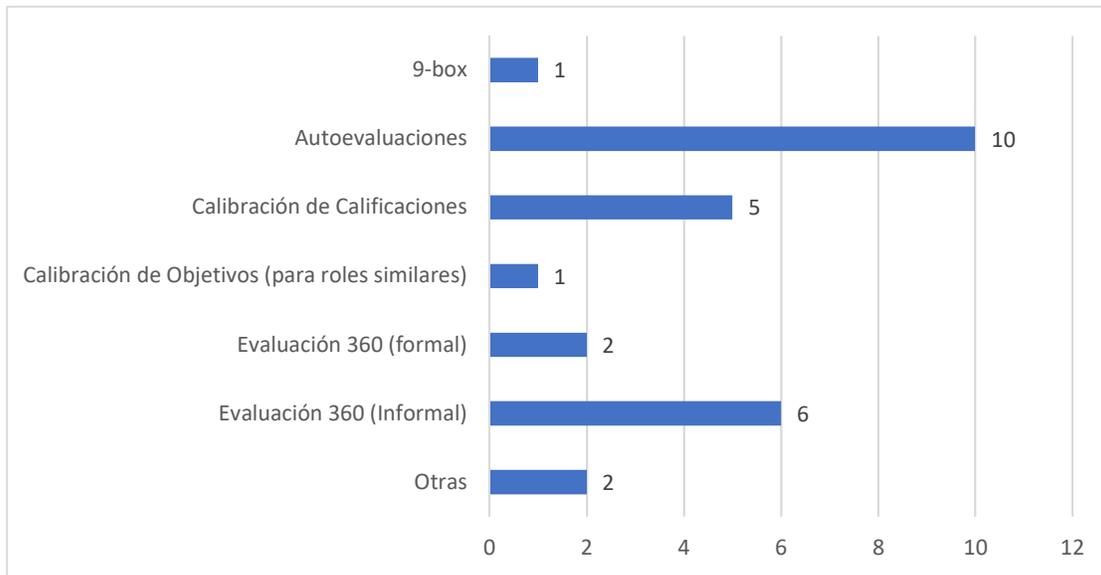


Figura 5 - Metodologías Usadas en su Herramienta de Evaluación

Resulta de interés la implementación de una empresa que se encontraba en proceso de actualización de su herramienta de evaluación del desempeño. A la fecha de la entrevista, estaban terminando de refinar la implementación de 9-box para el grueso de sus empleados y redefiniendo las competencias y calificaciones mínimas que deberían exhibirse para alcanzar una promoción. La calificación mínima para cada una de dichas competencias se volvía cada vez más exigente conforme el rol del evaluado conllevaba tareas más complejas y mayor responsabilidad. Del mismo modo, estaba en carpeta implementar un esquema de autoevaluación flexible por competencias, buscando que se incremente la participación de los evaluados en las reuniones de *feedback*.

Siguiendo con este caso, el plan continúa siendo que los líderes de la empresa (una vez que exceden cierta cantidad de empleados a cargo) completen el ejercicio formal de 360-*feedback*, involucrando especialmente a clientes internos y externos. Es un modo de reducir la demanda de tiempo y esfuerzo (que suele ser una de las principales contras de las evaluaciones 360): reservar la práctica solamente para posiciones estratégicas. Es evidente que la combinación de estas

diversas metodologías dentro de la empresa complejiza la implementación inicial, pero apunta a optimizar el nivel de esfuerzo requerido para obtener el nivel de detalle relevante para cada rol.

Otro caso de a destacar es el de la empresa que optó por abandonar el uso de herramientas de medición del desempeño implementadas y lideradas por su departamento de RRHH, y las reemplazó completamente por un seguimiento de la información provista por su implementación de Metodología Ágil. Más allá de definir y evaluar niveles mínimos de rendimiento para algunas de esas métricas, la implementación procura especialmente hacer foco en las áreas que cada empleado debe mejorar. Por ejemplo, aquellos que presentan desvíos muy marcados entre sus estimaciones y los tiempos reales de trabajo tendrán como objetivo achicar esa brecha, mientras que quienes deben mejorar la calidad de su código relevada a través de *peer-reviews* deberán trabajar en ese frente. Como resultado, la definición de objetivos reposa en observables cotidianos, que serán relevantes en tanto se provea *feedback* periódico. La estrategia al momento de la entrevista parecía estar resultando exitosa, aunque comenzaban a percibirse limitaciones a la hora de medir comportamientos no capturados diariamente por dichas metodologías (por ejemplo, habilidades blandas).

Finalmente, un respondiente declara que su líder apela pura y exclusivamente al método narrativo. Resulta entendible, dado que se trata de la empresa que abandonó el sistema de gestión de desempeño y está buscando el modo de sobreponerse a reducciones de personal a todo nivel. Apenas se busca asegurar un *feedback* mínimo que permita a cada evaluado identificar sus debilidades y fortalezas, prescindiendo de todo lo demás (incluso de la necesidad de calibrar calificaciones: se permiten notas idénticas para todos los miembros del equipo y esa decisión prácticamente no tiene impacto práctico debido a los mínimos ajustes de sueldo y nulas

promociones en los últimos 12 meses). En otras palabras, más que debido a la confianza en sus aportes, todo indica que esta metodología ha sido elegida por ser la única viable en el contexto.

En lo que respecta a garantizar resultados no uniformes, en la mayoría de los casos relevados (**11/14**) se confirma que no hay ningún método que fuerce los resultados esperados. Sólo **3** de las empresas tienen integrados a la herramienta métodos para asegurar que las calificaciones dentro de un equipo sean diferentes entre sus miembros. **1** de ellas apela a un lineamiento esperado de calificaciones que en la práctica termina funcionando de modo similar a una distribución forzada (aunque en ocasiones muy contadas se solicitan excepciones puntuales que rara vez son aprobadas por la dirección). Las otras **2** apelan al ordenamiento relativo de los empleados (tanto dentro de su equipo de desarrollo como contra otros equipos).

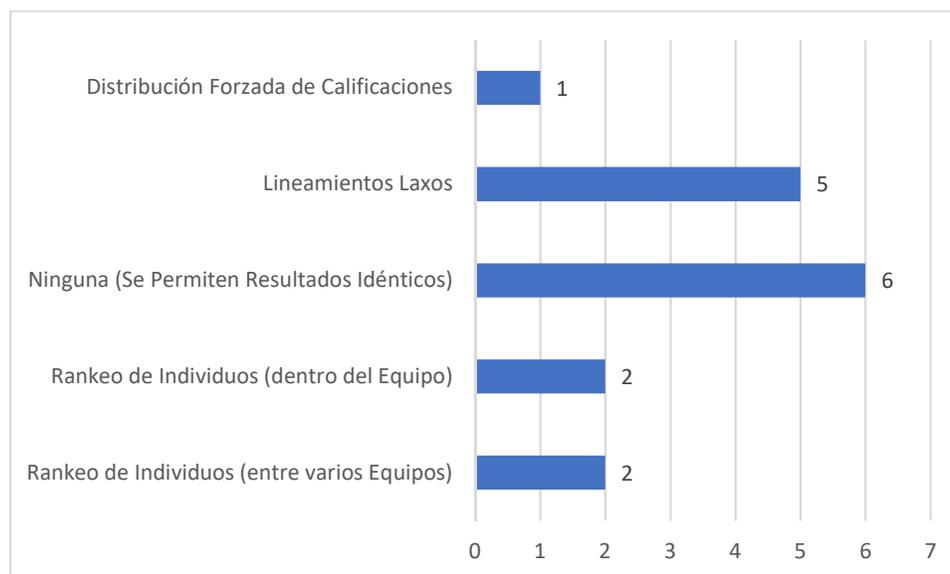


Figura 6 - Herramientas para Garantizar Dispersión de Calificaciones

De las **11** empresas que no fuerzan resultados esperados, **5** presentan lineamientos laxos sobre los objetivos o limitan la cantidad de recursos que pueden ser identificados como *key* (es decir, que pueden obtener la máxima calificación): por ejemplo, ponen como cota máxima al 20% del total de evaluados. Las restantes **6** no poseen mecanismos formales que fuercen calificaciones distintas.

Estos resultados guardan relación con las herramientas de calibración identificadas por los entrevistados. 4 de ellos declararon que no se utilizaba ninguna; en 9 de las restantes la calibración dependía de la revisión del jefe del evaluador. De este último subgrupo, en 3 se realizaba antes y después de las evaluaciones un ejercicio de calibración informal (se trata de aquella que limitaba la cantidad de *key performers* y las que perseguían el ordenamiento relativo de los empleados evaluados). Solamente en 1 empresa (justamente, la misma que tiende a hacer una distribución forzada) la calibración es también forzosa. Por último, en la empresa que usa una evaluación basada en el Método Narrativo no hay calificación que calibrar.

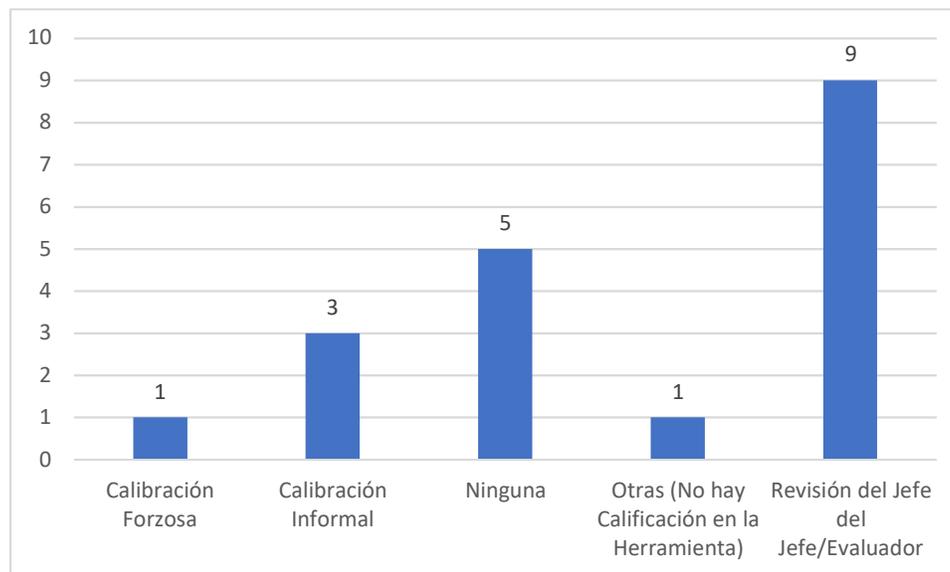


Figura 7 - Modos de Calibración

### III.C.III. Aplicación de los Resultados

En cuanto a los usos de la información compilada por las evaluaciones de desempeño, solamente 1 de las empresas relevadas se abstiene de tomarla como parámetro para tomar decisiones relacionadas con ajustes salariales (precisamente, la que apeló al Método Narrativo). Las 13 restantes sí fundamentan la remuneración en base a lo que se desprende de las evaluaciones de desempeño, lo cual constituye (por mucha ventaja) el uso más común.

En otro orden, y tal como se podía presuponer en una industria donde la competencia por el talento es feroz, **7** de esas **13** empresas optan también por otorgar ajustes salariales por potencial, reconociendo a los candidatos que se revelaban promisorios tras examinar sus competencias y contrastarlas contra las de sus pares de similar experiencia.

Continuando con formas de reconocer la performance positiva, **9** respondientes vinculan un determinado nivel de calificación a la hora de considerar promociones. Sorprendentemente, también son **9** las empresas que utilizan además la información para determinar el monto de pagos únicos (o bonos salariales por performance). Incluso considerando las reservas mencionadas en el Capítulo I, era esperable que fuera una práctica más extendida en la industria debido a lo competitivo del mercado. En contraste, ninguno de los respondientes afirma que se efectúan pagos únicos por potencial; ello sugiere que las retribuciones excepcionales todavía necesitan un fuerte correlato con resultados que excedan lo grupal.

En lo que respecta al desarrollo, **10** respondientes mencionan que el contenido de las evaluaciones de desempeño es tenido en cuenta a la hora de confeccionar planes de desarrollo individuales, inclusive a través de la creación de objetivos individuales y específicos a ese fin. **2** respondientes más aseveran que la evaluación de desempeño es utilizada para vincular al evaluado con planes de desarrollo predefinidos por necesidades estratégicas de la empresa. La suma de ambos enfoques permite inferir que las decisiones de desarrollo (con **12** respondientes) son la segunda aplicación más frecuente de la información recabada en las evaluaciones, solamente detrás de los ajustes salariales por performance. Esta decisión es coherente con la ya observada tendencia a hacer foco en objetivos y comportamientos, y la importancia que tiene el desarrollo personal tanto para estimular a los talentos de la organización como para generar en el interior de cada empresa los conocimientos y habilidades específicas de cada negocio o producto.

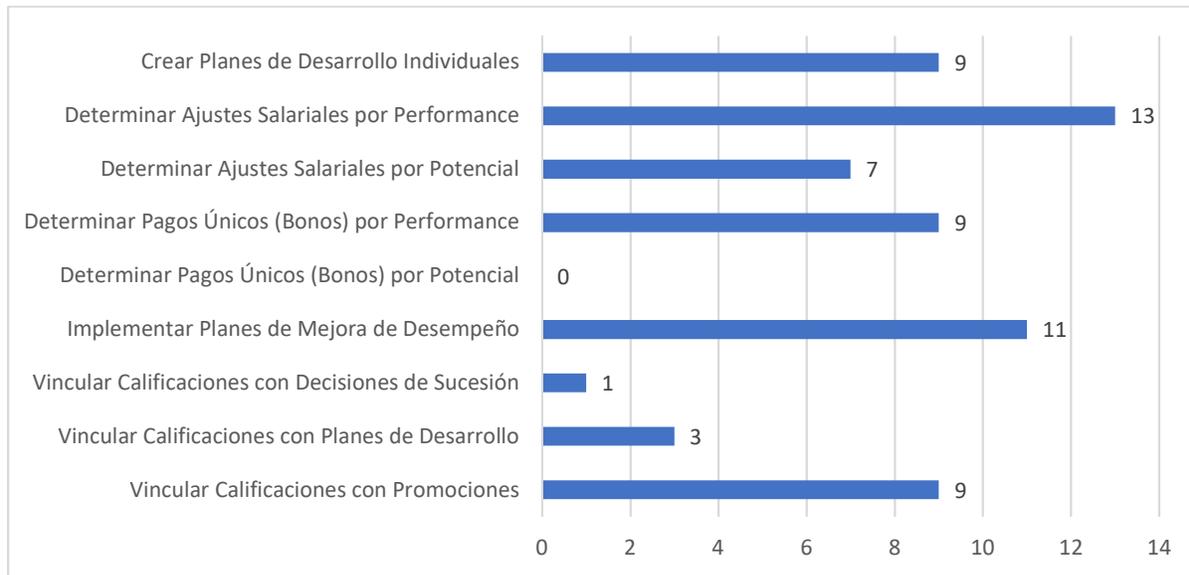


Figura 8 - Aplicaciones de la Información Individual Provista por la Evaluación de Desempeño

Como reverso de la anterior, **11** respondientes afirman que el contenido de las evaluaciones de desempeño puede determinar que los evaluados peor calificados y más alejados del cumplimiento de sus objetivos (o con dificultades para presentar los comportamientos deseados para su rol) sean introducidos en un programa de mejora del desempeño. Las características de estos programas varían mucho de empresa a empresa, según los objetivos de la herramienta de evaluación de desempeño, perfil del empleado en cuestión, decisiones del *management* y diversos factores adicionales.

Resulta destacable que las dos empresas que explícitamente utilizan la evaluación de desempeño para implementar planes de mejora del rendimiento (pero no la usan para programas de desarrollo) son justamente aquella que utiliza el Método Narrativo para evaluar a los empleados y la que establece una distribución forzada de calificaciones. Esto refuerza la noción de que la ausencia de desarrollo sólo deja espacio para un enfoque reactivo ante una baja performance, centrando la atención antes en la brecha de rendimiento que en los procesos que podrían estar

incidiendo en la performance del evaluado. Quizás esa reacción no sea el problema; sí es criticable la postergación permanente de la evaluación del desarrollo.

Volviendo a los usos menos extendidos, solamente 1 de las empresas parece utilizar la información vertida en evaluaciones de desempeño para tomar decisiones vinculadas a la sucesión de los líderes y directores actuales. Esto no quiere decir que en la práctica sea posible aspirar a ser promovido a un rol gerencial independientemente del contenido de las evaluaciones; antes bien, parece que las empresas que no usan el contenido de las evaluaciones para ese propósito no disponen de un criterio predefinido contra el cual se deba comparar el rendimiento de los evaluados. Así, en la mayoría de los casos los lineamientos a la hora de diseñar sucesiones únicamente comienzan a definirse al momento de empezar a diseñar la sucesión.

#### **III.C.IV. Participación de la Dirección**

Uno de los principales objetivos que una herramienta de evaluación de desempeño puede satisfacer es favorecer la propagación de la estrategia de la compañía. En este punto, cuanto mayor es la participación de la gerencia, tanto más probable es que la implementación de la herramienta sea exitosa: un alto involucramiento de la dirección implica mayores recursos, la autorización para dedicar tiempo (que generalmente es un bien escaso) y supervisión en cuanto al cumplimiento de plazos y objetivos.

Debido a que la muestra de respondientes fue confeccionada para satisfacer los propósitos de la presente investigación, fueron pocos los casos en los que se pudo acceder a la dirección para contar con respuestas de primera mano. No obstante, el cuestionario guía incluyó en todos los casos una pregunta que indagaba sobre el grado de involucramiento de la dirección en la construcción y aplicación de la herramienta de evaluación de desempeño.

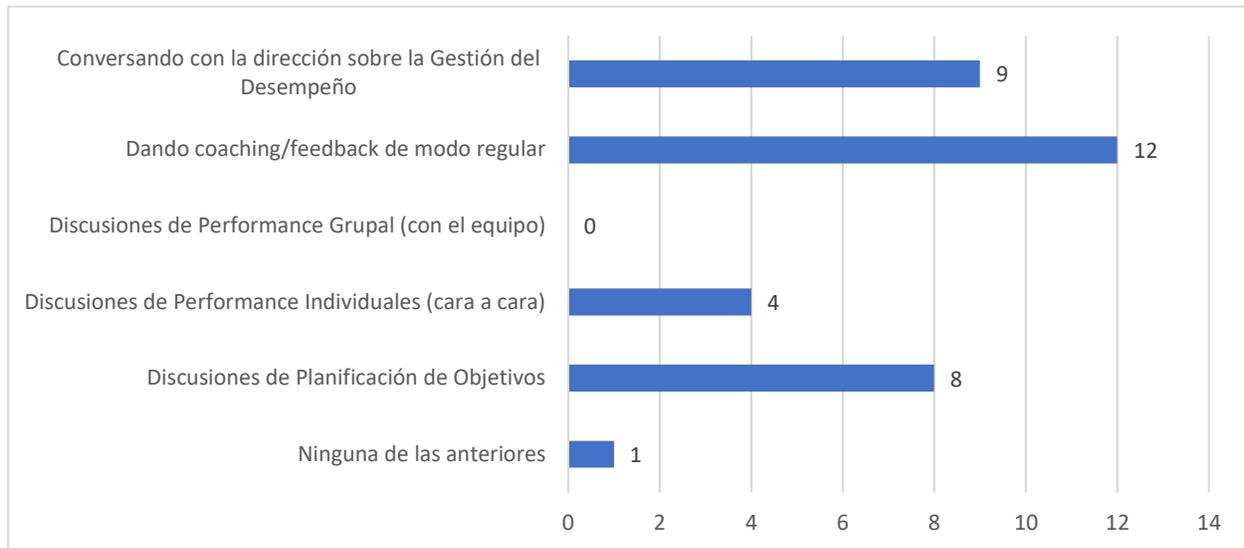


Figura 9 - Acciones en la que Participa la Dirección

Los resultados indican que, en los casos relevados, la dirección participa con aceptable intensidad en el soporte del equipo de gerentes y líderes a cargo de gestionar el desempeño. Solamente en una de las empresas (casualmente, aquella que cuenta con el sistema más rudimentario de evaluación) se percibe que el involucramiento de la dirección es nulo. En **12** casos la dirección participa dando coaching/feedback de forma regular, en **8** participa activamente en las discusiones de planificación de objetivos, en **9** discute con la gerencia sobre la estrategia de evaluación y su relación con el sistema de gestión del desempeño y en **4** se permite discutir cara a cara o evaluar el desempeño de sus gerentes/líderes.

A primera vista, las empresas donde la dirección realiza tres o cuatro de estas acciones a lo largo de cada ciclo de evaluación parecen tener poco en común: dos son consultoras y dos son empresas de producto; una persigue un negocio relativamente nuevo mientras que las otras tres trabajan sobre productos o servicios ya maduros; el tamaño de sus equipos de desarrollo y los planteles no son uniformes ni comparables en escala. Sin embargo, tras una segunda inspección un hilo conductor común recorre esas cuatro empresas: los fundadores aún tienen cargos directivos en las mismas y participan activamente en el diseño de la estrategia, en la propagación de la cultura

de la empresa y en la selección o evaluación de sus líderes y mandos medios. Se intuye que los fundadores/líderes, independientemente del tamaño que haya alcanzado la empresa, no delegan estos aspectos dado que les interesan a nivel personal o los consideran claves para el buen funcionamiento del negocio.

Desde ya, esto no quiere decir que la dirección profesional o de empresas globales tenga de por sí una desventaja. De hecho, uno de los casos relevados tiene una herramienta muy bien considerada por Managers y líderes que ha sido implementada por una dirección nueva. No obstante, sí puede verse una tendencia llamativa, que se comparte en este apartado adelantando un pequeño fragmento de las recomendaciones: el involucramiento de la dirección incide positivamente en el grado de satisfacción para con la herramienta que mide el desempeño. Para estos casos, ese nivel de satisfacción es decididamente mayor que el promedio en los casos analizados.

### **III.D. Metodologías Ágiles**

En ese apartado, se consolida la información provista por los entrevistados respecto a cuáles son las métricas que cimentan su análisis cotidiano del rendimiento del equipo. Dichas métricas están divididas en tres subgrupos: aquellas propias de las Metodologías Ágiles y el Negocio, aquellas relacionadas con desarrollo de software y aquellas relacionadas con calidad. Si bien en ocasiones la distinción entre estas entidades se torna difusa, ella adquiere relevancia en términos prácticos: las primeras impactan primeramente en los costos y fechas de entrega, la segunda está más relacionada con procesos internos o estándares de programación y productividad, mientras que en la tercera etapa se define si el producto es suficientemente satisfactorio para que lo revea el cliente o requiere retrabajo.

### III.D.I. Metodologías Ágiles: Una Advertencia

Como se recordará del marco teórico, las Metodologías Ágiles proveen mucha información valiosa para poder medir la velocidad del equipo a cada momento, la probabilidad de cumplir con las estimaciones para la próxima fecha de entrega y rentabilidades propias y del cliente. Es preciso resaltar que la mayoría de los respondientes mencionó que a la metodología de trabajo que ellos seguían se la podía llamar “Ágil”, *ma non troppo*: no era Metodología Ágil en el sentido estricto.

Algunas de las empresas relevadas han sido forzadas a espaciar sus entregables e incorporar un enorme volumen de trabajo para cumplir con una *Epic* al final de cada *Sprint* (lo cual resultaba en *Sprints* demasiado largos, y desvirtuaba las instancias de planificación). Otras parten de requerimientos detallados, lo cual transfigura la dinámica natural de dejarle al cliente la definición de *Acceptance Criteria* y librar al equipo para que decida sobre cómo alcanzar ese objetivo. Un tercer grupo explicita que no tienen ningún margen para renegociar fechas, con lo cual los desvíos en las estimaciones deben ser corregidos dentro del mismo *Sprint* (incluso cambiando su duración cada vez). Por último, un cuarto grupo de respondientes busca subrayar que no existen instancias de *Planning* o *Retrospectives*, o al menos no siguen los lineamientos recomendados, impidiéndole al equipo de desarrollo dar su *feedback* y debilitando sus estimaciones (o, peor aún, debiendo aceptar estimaciones impuestas por el cliente).

Es vital subrayar que cualquiera de estas acciones corrompe por completo a la metodología y (por ende) relativiza la relevancia de cualquiera de las métricas que se pueda recabar. ¿Qué sentido tiene medir el desvío respecto de una estimación que el equipo de desarrollo no hizo, no supervisó y que se vio forzado a aceptar? ¿Cómo puede estimarse confiablemente la velocidad futura del equipo si la información histórica de la que se dispone está basada en completar *Stories*

o *Epics* que llevan días o meses en vez de horas? ¿Qué mejora en los procesos puede adoptarse (y medirse) si no se planifica el trabajo ni se documentan los errores pasados o de qué modo se ha consensuado remediarlos?

En empresas que se dedican a la consultoría y tercerización de servicios de IT, el desafío es aún mayor. No solamente están obligados a lidiar con los desafíos anteriormente mencionados, sino que frecuentemente deben aprender a interpretar y comparar la información recabada en sus distintos equipos de desarrollo que siguen metodologías diferentes por indicación de sus clientes. La sintonía fina sobre los datos suele requerir una cantidad significativa de retrabajo, de traducción de las métricas recabadas o de discusiones con proveedores (o clientes internos y externos) hasta lograr recabar suficiente contexto para poder interpretar la información con algo de confianza.

Similares obstáculos se presentan en casos en los que la actividad que se rige por esas metodologías tiene otra naturaleza que el desarrollo de software. No es extraño que haya recursos especializados en automatización de pruebas, *data science* y *business intelligence*, por ejemplo, desempeñándose dentro de un equipo de desarrollo. Ello requiere encontrar un modo de adoptar la misma metodología que sus pares incluso cuando sus objetivos son diferentes, la dinámica diaria tiene poco en común, sus avances suelen darse a un ritmo más espaciado y los criterios de éxito o fracaso son infinitamente más debatibles.<sup>33</sup>

Alentar a los equipos y clientes a que adopten Metodologías Ágiles requiere explicar sus beneficios y cómo ellos compensarán los desafíos, incluyendo la imposibilidad de seguir métodos

---

<sup>33</sup> Un entrevistado subraya esta tensión del siguiente modo: “en mi caso, muchas de esas métricas no aplican. Mi trabajo puede haber sido impecable y se lo puede reconocer incluso si las conclusiones de mi análisis no validan la hipótesis previa. OK, mi teoría no se verifica en los datos, pero eso es información valiosa que antes no teníamos: sirve aunque no haya funcionado. Para un programador, en cambio, algo así en una implementación es un problema enorme”.

de negocio tradicionales (como conocer el costo total antes de empezar el proyecto), la necesidad de entregar decisiones técnicas y de diseño al equipo de desarrollo (en vez de tener la última palabra sobre el funcionamiento del producto) y la aceptación de las estimaciones de dicho equipo (en vez de forzarlos a aceptar las propias). Desde ya, varias de las empresas entrevistadas para este trabajo lo han intentado sin éxito y han debido optar entre perder un negocio y sacrificar pilares esenciales de las metodologías.

Agotar las razones por las cuales se produce este fenómeno excede las capacidades del presente trabajo. Pese a ello, se los menciona en este apartado con dos propósitos. El primero, matizar a la luz de estas limitaciones la calidad de las métricas que dicen relevarse. El segundo, mencionar las razones por las cuales la teoría sobre las Metodologías Ágiles presenta supuestos que tienen muchas barreras para poder ser llevadas a la práctica.

### **III.D.II. Métricas Relacionadas con la Metodología y el Negocio**

Como se recordará tras lo expuesto en el Capítulo II, algunas de estas métricas solamente adquieren relevancia cuando son analizadas a nivel grupal. Todas ellas fueron mencionadas entre las que se recaban y se analizan por parte de todos los entrevistados. Las más populares son el cumplimiento de *Schedule Forecast* (es decir, acatar los compromisos respecto de fechas y presupuesto), relevada por los **14** entrevistados, seguida del Análisis del *Burn-Down chart* adoptado por **10** respondientes. Estas métricas tienen la enorme ventaja de su simplicidad y su relevancia: haber cumplido con las fechas estimadas es un buen modo de calcular si se ha cumplido con los objetivos del proyecto dentro de los plazos estimados, mientras que el análisis del trabajo pendiente permite detectar desvíos de forma temprana (y, por ende, resolverlos antes de que amenacen la fecha de finalización del proyecto). Son además relativamente fáciles de obtener y

analizar (de hecho, casi que se desprenden de los progresos diarios). Esto baja enormemente el costo de implementarlas.

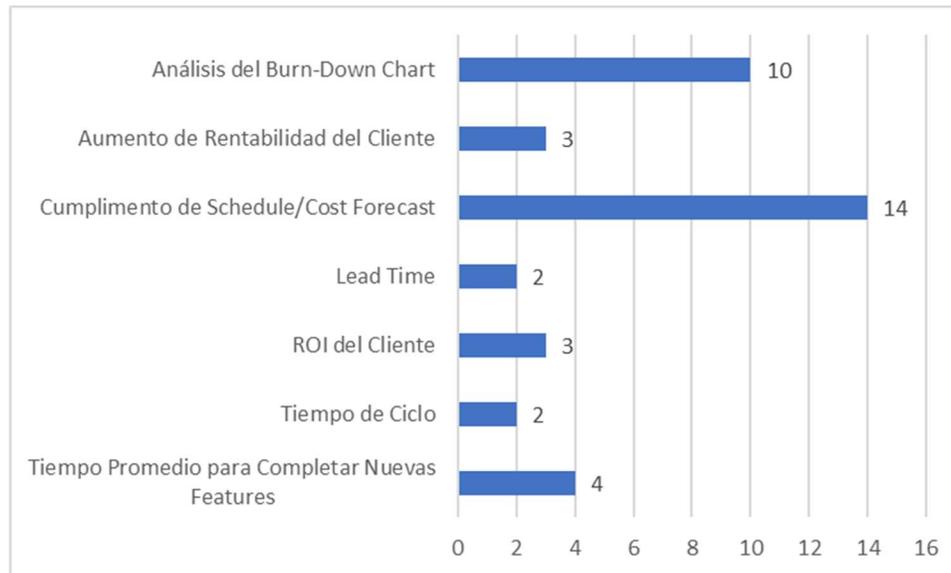


Figura 10 - Métricas Relacionadas con la Metodología y el Negocio (Sólo Grupales)

Las demás métricas son utilizadas de forma dispar por las diferentes empresas. No obstante, es para resaltar que solamente **5** de los respondientes afirman que el análisis se circunscribe a las dos métricas mencionadas en el párrafo precedente. En otras palabras, las restantes **9** empresas enriquecen el análisis de avance diario y fechas de entrega con las demás métricas (más granulares). **4** de ellas buscan calcular el tiempo promedio para entregar una nueva *feature*, **3** buscan mensurar el aumento de la rentabilidad del cliente (una vez terminada una implementación) y **2** empresas (no necesariamente las mismas para todas las métricas) optan por capturar el Tiempo de Ciclo, el Lead Time y el ROI del Cliente.

No parece haber más motivos para la elección de una u otra métrica que la colaboración del cliente o el impacto de los nuevos entregables. Por ejemplo, para una empresa especializada en producir aplicaciones para dispositivos móviles es bastante sencillo estimar cuántos usuarios finales se descargan una aplicación nueva y cuántos de ese grupo optan por pagar una suscripción,

con lo cual la monetización del trabajo es fácil de obtener y calcular. Empresas que proveen un producto y lo adaptan a las solicitudes de clientes puntuales seguramente obtendrán más beneficios al conocer de antemano qué tiempo se estima para cada nueva solicitud de implementación, y por tanto les resultará más relevante compilar información histórica para calcular ese dato.

Dentro de este apartado se incluyó también una pregunta orientada a analizar la frecuencia con la que se recaba información que puede ser analizada tanto a nivel individual como grupal. Se esperaba que los entrevistados refirieran un uso relativamente uniforme de estas métricas; en otras palabras, que cuando el análisis de una métrica fuera concebido como valioso a nivel grupal, se buscara procesarla para imputarla también a nivel individual. No fue el caso. **12** respondientes mencionan que se analiza grupalmente el desvío entre las estimaciones y el tiempo efectivo en implementar una tarea, pero solamente en **4** casos (generalmente, cuando los desvíos eran consistentes y de gran magnitud) se busca vincular esa información a nivel individual. Algo similar ocurre con la medición subjetiva (y generalmente informal) de la satisfacción del cliente para con el equipo de trabajo: **13** empresas capturan esa información a nivel grupal, pero solamente en dos casos se consolida dicho *feedback* a nivel individual.

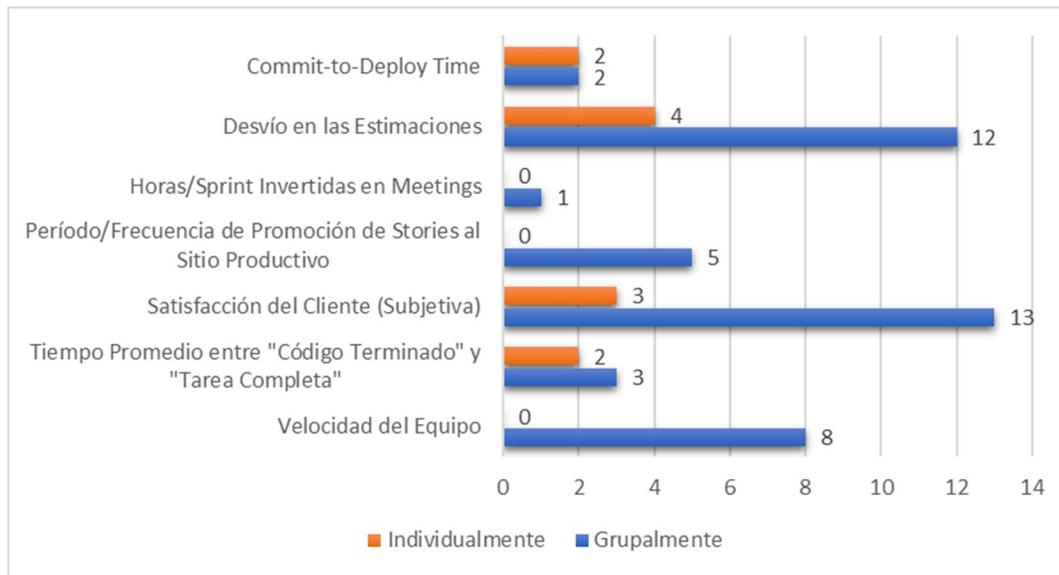


Figura 11 - Métricas Relacionadas con la Metodología y el Negocio (Individuales/Grupales)

Estas conclusiones tienen como excepción a las dos métricas relacionadas con la velocidad entre que una *Story* es finalmente implementada (*Commit-to-deploy time*) y Tiempo Promedio entre estados “Código Terminado” y “Tarea Completa”. La explicación intuitiva es que, a diferencia de las dos métricas mencionadas en el párrafo anterior (que frecuentemente están disponibles a nivel grupal), en las últimas se parte generalmente de información individual y es necesario agregarla manualmente. De tal modo, es fácil analizarla a ambos niveles. Eso permite comprender por qué **2** empresas procesan ambas métricas de modo individual y grupal. Únicamente **1** empresa opta por capturar la métrica de “Tiempo Promedio...” sólo a nivel grupal, con el fundamento de que sus *Sprints* son especialmente largos y por lo tanto el esfuerzo invertido en control de calidad (el cual esta métrica captura indirectamente) puede amenazar el alcance de una fecha de entrega.

Las demás métricas capturadas en este apartado son procesadas únicamente a nivel grupal. Se trata de Horas/Sprint invertidas en reuniones (solamente considerada relevante por **1**

respondiente), Período/Frecuencia con el que se promueven cambios a Producción (capturada por 4 empresas) y *Sprint Velocity* (la tercera más popular, siendo utilizada por 8 empresas).

### III.D.III. Métricas Relacionadas con Web/App Development

En este grupo de preguntas, se busca relevar la adopción de varias de las métricas de desarrollo de software abordadas en el marco teórico, tanto a nivel grupal como individual. Tal como se suponía, dichas métricas son conocidas por la totalidad de los respondientes. La única excepción fue la referida a Eficiencia del Código (o su reverso, tasa de código improductivo), cuya definición no estaba del todo clara para uno de los entrevistados, pero fue reconocida una vez definida.

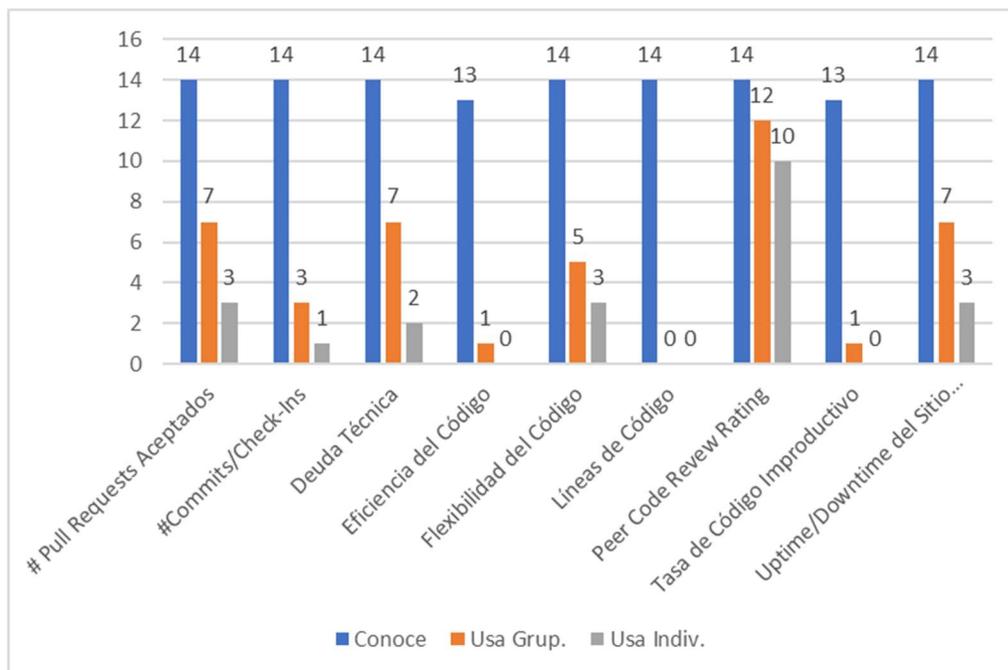


Figura 12 - Métricas Relacionadas con Web/App Development

La métrica más utilizada (tanto a nivel grupal como individual) es la de *Peer Code Review rating*. 12 respondientes declaran que se la usa a nivel grupal, mientras que 10 la utilizan a nivel individual. Como una de las firmas la usa exclusivamente a nivel individual, en total 13 de los 14 casos relevados utilizan esta práctica (y consolidan la información recabada en el proceso) para

evaluar la robustez del código producido. Esto confirma que, aunque el esfuerzo invertido en esta práctica es alto (dado que es difícil de reemplazar por procesos automatizados), es el mejor modo de asegurar el seguimiento de patrones de desarrollo, evaluar el cumplimiento de los objetivos y (a largo plazo) la calidad y mantenibilidad del código. Aquellos que utilizan esta práctica a nivel individual, buscan especialmente hacerlo para mejorar la adopción de estándares de programación y resolver errores de concepto, aunque en un caso la práctica fue llevada a cabo para buscar evidencia empírica para sustentar una promoción.

En el otro extremo, se ubica la métrica de Líneas de Código. Conocida por todos los entrevistados, ninguno de ellos opta por capturarla dado que su relevancia es baja (es decir, para obtener información a partir de ella se la debe entrecruzar con varias de las otras métricas posibles). A ella le siguen las métricas sobre Eficiencia del Código y Tasa de Código improductivo, solamente utilizadas a nivel grupal por uno de los respondientes. Esta última métrica requiere un esfuerzo moderado para ser recabada a largo plazo y robustecerla con análisis (a fin de poder determinar, por ejemplo, si el código que se ha quitado efectivamente fue removido por ser redundante o ineficaz o, más bien, porque hubo un cambio de requerimientos o un rediseño de la solución que excedió a la voluntad del programador). Esto explica que solamente haya sido adoptada por una empresa donde los patrones de diseño son muy estrictos, la capacitación a los programadores es constante y la promoción del código a instancias productivas solamente ocurre luego de controles muy rigurosos. En una empresa de este estilo, buena parte del trabajo de análisis está hecho de todos modos, con lo cual la información puede ser interpretada con bajo esfuerzo.

Luego, las empresas de producto son las que con más frecuencia toman métricas que se referían a la disponibilidad del ambiente de producción (*Uptime/Downtime*) y Deuda Técnica: se trata de 7 respondientes que capturan cada una de esas métricas a nivel grupal (subrayamos que no

siempre esas empresas llevan cuenta de ambas). **2** de las empresas profundizan ese análisis hasta poder vincular individualmente a un programador con los resultados observados. También son las que con más frecuencia analizan la calidad y flexibilidad del código (con **5** adopciones grupales y **3** interpretaciones a nivel individual), siendo esta última un apartado específico de las *Peer Code Reviews*.

Al igual que las prácticas de *Peer Review*, Deuda Técnica y Flexibilidad del Código son métricas que requieren de análisis de recursos especializados. Las métricas que cuantifican la disponibilidad de los ambientes de producción son relativamente sencillas de capturar en comparación, aunque debe resaltarse que los responsables de producir y actualizar el código de esos servidores son quienes deben determinar su relevancia. Por ese motivo, es razonable que solamente las empresas que son responsables tanto del mantenimiento como del crecimiento de un producto presten especial atención a estos tres aspectos. Por un lado, debido al costo inmediato (tanto económico como en percepción de calidad, de servicio) de una interrupción en el funcionamiento de los servidores; por otro, por el costo futuro de adquirir deuda técnica o producir código inflexible o difícil de escalar.

Finalmente, una menor cantidad de las empresas opta por llevar registro de métricas referidas a la cantidad de *Check-Ins* (capturadas por **3** respondientes a nivel grupal y **1** a nivel individual) o ratio de *Pull Requests* aceptados (**7** a nivel grupal y **3** a nivel individual). La diferencia de adopción entre ambas obedece a que la mera captura de cantidad de “*Check Ins*” tiene limitaciones analíticas similares a las de cantidad de líneas de código: sin análisis agregado, es una métrica que dice poco. En cambio, la aceptación de *Pull Requests* sigue procesos comparables al de *Peer Code Reviews* – quizás con menor profundidad-. En cualquier caso, estas métricas ofrecen conclusiones que sirven de complemento a las anteriormente mencionadas (y capturadas con más

frecuencia). Están, por lo tanto, presentes solamente en empresas que tienen una alta madurez en lo que respecta a medir cuantitativamente sus progresos.

### III.D.IV. Métricas Relacionadas con Calidad del Código

En lo que respecta a este tipo de métricas, la distribución parece arrojar mayor coincidencia entre los respondientes. Para empezar por el punto de consenso, los **14** sistemas de evaluación le dan trascendencia grupal a la cantidad de defectos encontrados una vez que el código ya ha sido promovido al sitio productivo (o, en otras palabras, a la cantidad de defectos encontrados cuando ya es demasiado tarde, lo cual habla negativamente tanto de la calidad del código como del proceso de control de calidad). Solamente **1** de ellas analiza la causa raíz del incidente hasta identificar a (o a los) responsables del error individualmente. Esta última práctica debería evitarse siguiendo los preceptos de la metodología que nos interesa estudiar; hemos mencionado anteriormente que todas las prácticas y cada uno de los resultados son responsabilidad del equipo como un todo, lo cual desalienta la identificación individual de las responsabilidades.

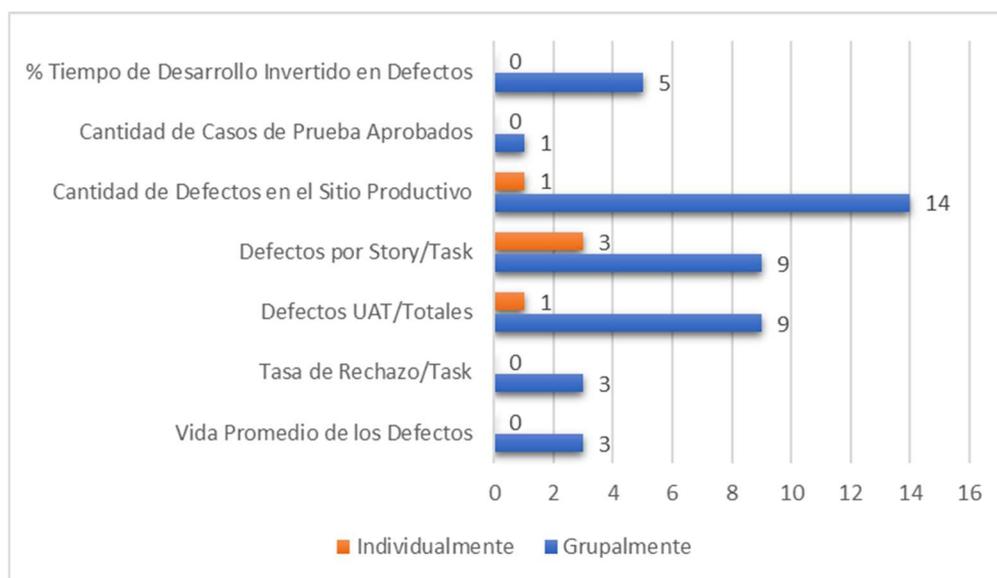


Figura 13 - Métricas Relacionadas con la Calidad del Código

No obstante, lejos de un ejercicio de castigo, ese único caso de individualización obedece a la necesidad de asignar la tarea para remediarlo. En esa empresa, los programadores suelen dividirse tareas por áreas de competencia. En tanto no exista una infrecuente recurrencia de defectos atribuibles al trabajo de un recurso puntual, ni siquiera se relaciona estos episodios con programas de mejora del desempeño. Es, por lo tanto, una práctica razonable en el contexto, merced a su necesidad práctica.

En términos de frecuencia de ocurrencias, detrás de la métrica anterior se encuentra la que vincula aquella que relaciona los defectos encontrados para cada *Story/Task*. Más allá de que sigue siendo una métrica predominantemente grupal (para **9** respondientes), la dinámica identifica responsabilidades individuales en **3** firmas. La mayoría de los respondientes subraya que esta métrica grupal solamente adquiere relevancia en caso de que la distribución de defectos se concentre en una determinada tarea sin que medie una alta complejidad que lo justifique. Cuando se perciben demasiadas fallas para la relativa sencillez de la tarea, las principales hipótesis son que se haya entendido incorrectamente el criterio de aceptación o que el código producido sea de pobre calidad.

Teniendo presente ese objetivo grupal, es razonable que algunos de los respondientes procuren llevar ese análisis hasta el nivel individual cuando se dan estas condiciones. De ese modo, la amplia diferencia entre la cantidad de respondientes afirmativos para el nivel grupal e individual es atribuible a incidentes recientes que requirieron análisis y trabajo al nivel mencionado.

Siguiendo criterios idénticos a los de la cantidad de defectos en el sitio productivo, se encuentra la métrica de cantidad de defectos encontrados durante el proceso de UAT (también llamado “testeo de aceptación por parte del usuario final”). La misma es recabada por **9** empresas

grupalmente y nada más que **1** a nivel individual (como es esperable, se trata del mismo respondiente que la capturaba para defectos en el sitio productivo, y por las mismas razones). Se postula que son métricas de naturaleza similar porque encontrar defectos en esta etapa del proceso es también hacerlo demasiado tarde; la diferencia es que el impacto sobre clientes y usuarios finales es mucho más reducido, precisamente porque se produce antes de que el código se promueva al sitio productivo.

Ahora bien, ¿por qué si la lógica es similar hay diferencias palpables en el grado de adopción de esta métrica? Según las opiniones recabadas, esto obedece a que no todos los entrevistados siguen procesos que distingan formalmente las pruebas de aceptación del usuario de la instancia formal de pruebas del equipo de control de calidad. Dicho de otro modo, no se trata de que todos los entrevistados sigan el proceso y ponderen la métrica de modos diferentes; antes bien, en aquellas donde la dinámica de UAT es comparable y se la cumplimenta formas análogas a las de otros respondientes, el valor y la aplicación de la métrica son similares.

Las demás métricas sobre la calidad del código tienen menor nivel de adopción. Cuando la hay, es exclusivamente de modo grupal. De mayor a menor adopción están el Porcentaje del Tiempo Total de Desarrollo dedicado a reparar defectos (5 empresas), la Vida Promedio de los Defectos y Tasa de Rechazo (3 empresas cada una) y por último la Cantidad de Casos de Prueba Aprobados para cada *Task* (1 empresa). De modo análogo a las métricas de desarrollo, estas últimas métricas no son complejas de recabar, pero cada una de ellas interpretada de modo aislado dice poco si no se la relaciona con *Tasks/Stories* o no se las vincula con la cantidad de defectos que son encontrados en una instancia demasiado tardía. Por ende, son métricas que solamente son capturadas por empresas cuyo proceso de revisión y desarrollo ha alcanzado un nivel de madurez

tal que las ofrece sin que ello agregue esfuerzo, y pueden por lo tanto echar mano a la posibilidad de enriquecer el contexto con su análisis sin pagar un costo relevante por ese aporte.

### **III.D.V. Reuniones Retrospectivas: ¿Cómo se consolida esa información?**

El cuestionario documenta no solamente qué información se recababa en las reuniones retrospectivas y con qué frecuencia; también indaga sobre el modo en que se consolidan los datos y se los integra con las evaluaciones de desempeño.

Sorprendentemente, son mayoría los respondientes que sostienen que no celebran reuniones retrospectivas después de cada ciclo ni hay instancias formales de revisión y ajustes de los procesos. Hay muchos modos de alejarse del método recomendado: están quienes sostienen que la reunión no existe (y el *feedback* es solamente discutido oralmente entre pares), aquellos que aseveran que solamente circulan notas aisladas por correo electrónico, los que afirman que cada equipo se administra según su propio criterio (con lo cual los lineamientos solamente se consolidan si ha habido una situación verdaderamente crítica) y quienes dicen que apenas se compartían notas en Hojas de Cálculo que no vuelven a ser revisadas una vez terminado el ejercicio.

Quienes procuran seguir el ejercicio recomendado, lo hacen de modo informal y con documentación escasa. Como regla general, se registran puntos a mejorar grupalmente por correo electrónico, hojas de cálculo, sistemas internos de cada proyecto o mensajería instantánea. Pasado ese punto, el seguimiento queda librado al esmero del grupo o el celo de su líder. Sólo 3 respondientes afirman que el ejercicio se realiza con frecuencia regular y se documenta en repositorios o herramientas especiales para ese propósito.

Si el ejercicio de revisar procesos y consolidar elementos de mejora no es llevado adelante, no puede sorprender que las lecciones aprendidas tampoco sean integradas a las evaluaciones de

desempeño en la mayoría de los casos. En ocasiones, los líderes de equipo son suficientemente puntillosos como para sobreponerse a las limitaciones del proceso y hacer un seguimiento de correos electrónicos para que sus conclusiones y recomendaciones llegaran a asentarse en las evaluaciones de desempeño, pero la consolidación de la información es de todas maneras discrecional. Solamente 2 de los respondientes declaran que las herramientas de medición de desempeño han sido concebidas con categorías específicas para este tipo de información: en la primera de ellas, la categoría donde se registran premios vinculados a mejoras de procesos es “resolución de problemas”, mientras que en la segunda se busca relacionar esos avances con las competencias de cada programador.

Aquí se revela una de las principales limitaciones impuestas por las dinámicas de las metodologías ágiles implementadas de modo imperfecto o ad-hoc para cada cliente. Si bien hay muchos motivos válidos para recolectar métricas a nivel grupal (y los hemos mencionado en esta sección), las Metodologías Ágiles requieren hacer foco en cuestiones de proceso que no siempre son sencillas de monitorear (y mucho menos de modificar). Si se acepta que resolver problemas grupales requiere introducir cambios en procesos compartidos, probablemente se concuerde con que uno de los modos más eficaces de monitorear mejoras a ese nivel es vincular la contribución individual a la mejora de los procesos grupales. La evaluación de desempeño (especialmente cuando se la realiza con suficiente frecuencia) debería ser un aliado a la hora de mensurar estos aspectos, tanto para volverla más representativa como para motivar a los empleados a aplicarse en una tarea cuyos beneficios no siempre son evidentes ni inmediatos.

### **III.E. Percepción General del Sistema**

Esta última subsección busca obtener de cada respondiente su percepción sobre la eficacia de la herramienta de evaluación de desempeño para diferentes objetivos, y una impresión definitiva sobre la centralidad de dicha herramienta a la hora de determinar aspectos cruciales como ajustes de sueldo, recategorizaciones y planes de desarrollo. El anhelo es detectar no solamente la opinión sobre la herramienta de modo aislado, sino contrastar esa opinión con su relevancia práctica. No es tan grave que una herramienta sea muy mejorable; sí lo es que, incluso sabiendo que tiene defectos claros, se la utilice como base para tomar decisiones equivocadas.

#### **III.E.I. Grados de Satisfacción con la Herramienta Actual**

La distribución de respuestas sobre la eficiencia de las herramientas para medir el alcance de objetivos individuales es bastante dispar, aunque de distribución celosamente normalizada. El promedio de las respuestas recae en que el sistema es aceptable, con **4** respondientes (28.57%); **3** eligen tanto la calificación “buena” como “regular” (21.42%), y **2** más eligen “mala” y “muy buena” (14.28%). Por supuesto, las empresas que están más lejos de una implementación de una herramienta que gestiona por objetivos y comportamientos son las que manifestaron más insatisfacción, y los **5** respondientes más conformes trabajan con formularios en cuya creación han participado y que ha madurado con el correr del tiempo hasta llegar al punto en que miden lo que efectivamente es valioso y con las herramientas adecuadas.

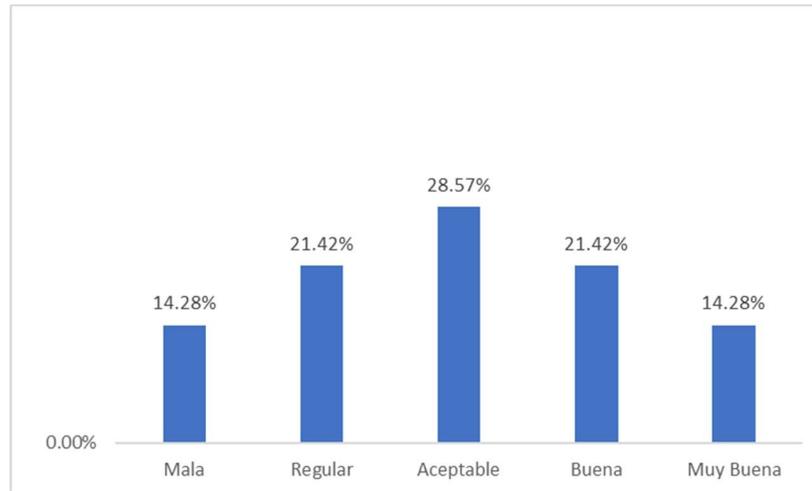


Figura 14 – Percepción de la Eficiencia con que la Herramienta de Evaluación mide Objetivos Individuales

Menos uniforme (y más sorprendente) es que la mayoría de los respondientes estima como regular o mala la capacidad de la herramienta de evaluación de desempeño para medir potencial: **5** (35,71%) la califican como mala, mientras que **6** (42,86%) como regular. Tan solamente **2** la califican como aceptable (14,29%), y sólo **1** respondiente como buena (7,14%). Este factor puede no tener consecuencias prácticas inmediatas, y es entendible que las herramientas de evaluación de desempeño que todavía buscan medir eficazmente los objetivos tengan limitaciones para identificar candidatos con alta capacidad de desarrollo. No obstante, en un mercado tan competitivo como el de Desarrollo de Software, documentar y reconocer el potencial es un aporte crucial para la gestión del talento.

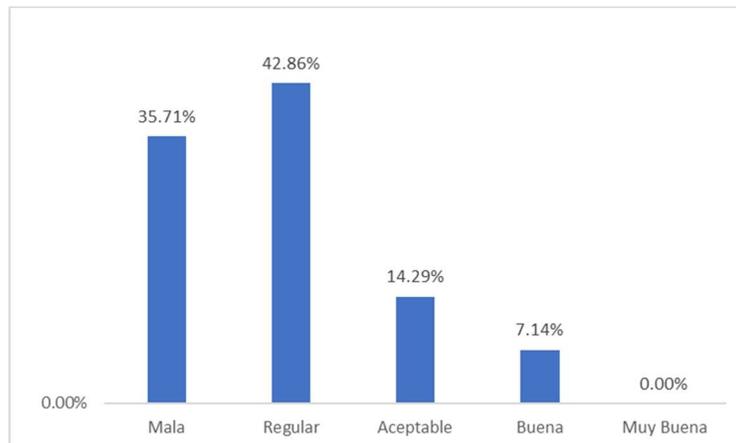


Figura 15 - Percepción de la Eficiencia con que la Herramienta de Evaluación mide Potencial Individual

En general, los respondientes aseveran que exceder los objetivos es una aproximación decente a la identificación del alto potencial, o bien que todos los miembros de un equipo saben identificar a ese tipo de recursos. En ese caso, quizás solamente es necesario documentarlo adecuadamente o diseñar indicadores que permitan identificarlos de modo más objetivo; en cualquiera de los casos, parece ser un área donde las herramientas (incluso las mejor consideradas por los entrevistados) tienen oportunidades de mejora.

El reverso de lo explicitado en el párrafo anterior ocurre a la hora de medir la colaboración de cada individuo con su equipo de trabajo: la opinión es mayormente positiva. **2** respondientes califican a la herramienta en uso como muy buena a esos fines, mientras que **8** la califican como buena (57.14%) y **1** como aceptable (7.14%). Solamente **1** la considera regular y **2** (14.29%) como mala. Estos resultados obedecen a que las empresas se han tomado el tiempo de integrar habilidades interpersonales y colaboración a la mayoría de sus herramientas, de modo tal que los recursos sólidos en lo técnico pero que trabajan aislados no tengan un reconocimiento desmedido. De hecho, de los **3** respondientes que no están conformes con la eficacia de la herramienta para este objetivo, **2** declaran que se mide de un modo con el que no estaban de acuerdo, con lo cual la

observación no se centraba en una omisión grave en lo que refiere a este aspecto, sino en la ineficacia de la operacionalización de los observables.

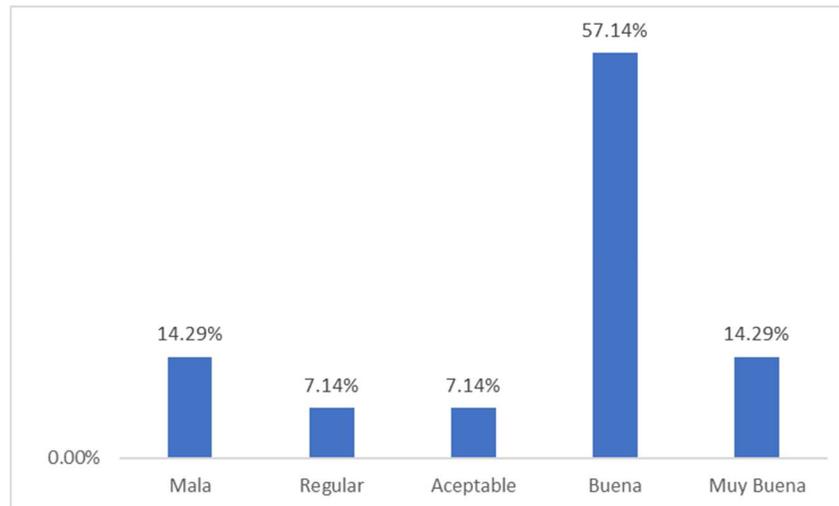


Figura 16 - Percepción de la Eficiencia con que la Herramienta de Evaluación mide Colaboración con el Equipo

Por último, quienes se manifiestan satisfechos con los aportes de su Herramienta de Evaluación de Desempeño en los tres lineamientos mencionados amplían su respuesta del siguiente modo: aseveran que lo que ellas miden es relevante para la estrategia de la empresa, opinan que es una de las razones por las que está mejorando su rentabilidad, perciben a la herramienta como una aliada para el desarrollo personal y la entienden como un paso más hacia la adopción de la cultura de la empresa. Como se observa, son opiniones alineadas con los objetivos de este tipo de sistemas de evaluación que se han mencionado en el Capítulo I del presente trabajo.

### III.E.II. Peso Práctico

En estos apartados se ve una fuerte correlación (como era de esperarse) entre el grado de satisfacción con la herramienta y el peso práctico que se le asigna para cada utilidad real. Así, herramientas muy pobremente calificadas a la hora de mensurar el cumplimiento de objetivos individuales tienden a no ser percibidas como relevantes a la hora de considerar ajustes salariales

o promociones. Con frecuencia, se las sustituye por la dinámica e información cotidiana que proveen las Metodologías Ágiles de desarrollo de software. En circunstancias como esa, se revela que la evaluación de desempeño se vacía de sentido, convirtiéndose ya sea en un proceso burocrático y rutinario que va por carriles diferentes a aquellos donde se consolida la información relevante para la empresa, o bien en un proceso donde el líder de un equipo se limita a documentar lo que sea necesario para refrendar su impresión sobre un empleado.

8 respondientes (57%) dicen que la información vertida en la evaluación de desempeño es la que tiene preeminencia a la hora de considerar recategorizaciones; es una diferencia algo estrecha con los 6 que opinan lo contrario (43%). Justamente, en empresas donde las recategorizaciones requieren una revisión exhaustiva de las últimas evaluaciones de desempeño es donde la herramienta tiene más chances de ser relevante a ese fin. Las demás empresas consideran recategorizaciones por otros mecanismos vinculados a metodologías ágiles u otros criterios de seguimiento cotidiano, ya sea porque es preciso tomar rápidamente una decisión para llenar una vacante o porque los procesos no demandan demasiado trabajo adicional.

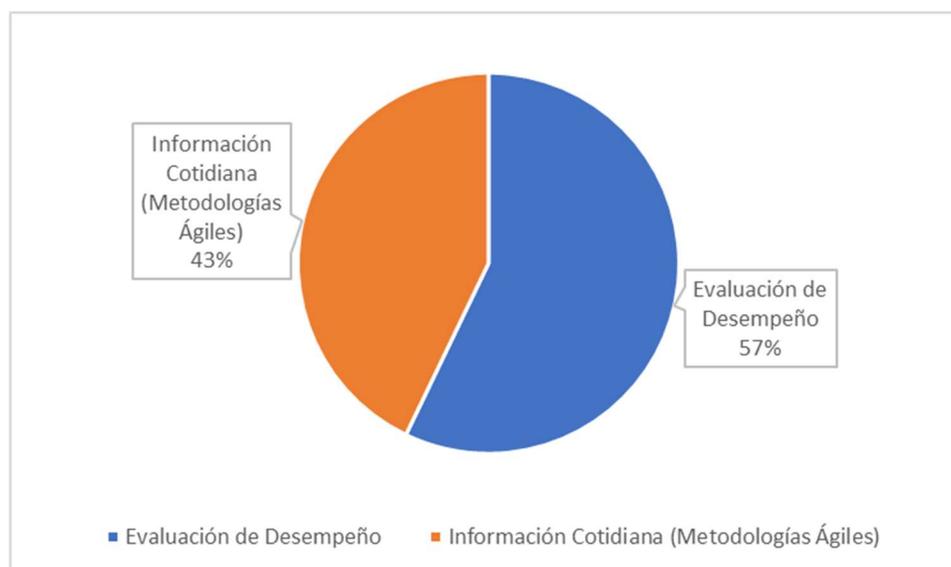


Figura 17 - Percepción de Mayor Peso Práctico (Recategorizaciones Individuales)

Los respondientes que declaran que las evaluaciones de desempeño tienen preeminencia a la hora de decidir retribuciones salariales individuales crecen hasta **10** de los **14** casos estudiados (71%). Esto confirma la opinión del 100% de los entrevistados respecto a que las evaluaciones de desempeño son utilizadas como información clave a la hora de determinar ajustes de sueldo. En otras palabras, servir de base para tomar esta decisión puntual parece justificar en muchos casos la propia existencia de la herramienta, lo cual a su vez realimenta su peso práctico en este apartado.

Es más, los **4** respondientes (29%) que dijeron que la información cotidiana tenía más peso no niegan que la evaluación de desempeño tenga incidencia; antes bien, declaran que las decisiones salariales dejan lugar para discrecionalidades (por ejemplo, ajustes de sueldo fuera de plazo o por valores por encima de los de un grupo) para casos excepcionales donde los líderes de un equipo identifican como clave a un empleado insatisfecho y cuya evaluación de desempeño no justifica un ajuste de ese nivel. Por lo tanto, las respuestas a esta pregunta (aunque a primera vista subraye la utilidad de la herramienta) parecen señalar más una debilidad que una virtud: que se vincule a las evaluaciones de desempeño principalmente con este aspecto habla de lo mucho que se pueden mejorar sus prestaciones en términos estratégicos y de generación de valor.

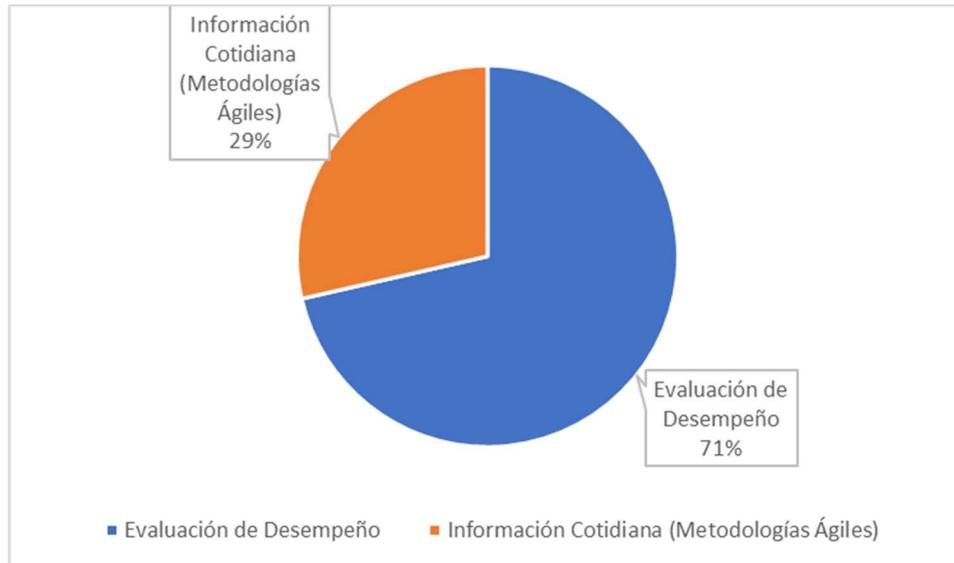


Figura 18 - Percepción de Mayor Peso Práctico (Retribución Salarial Individual)

Por último, a diferencia de los casos anteriores, los respondientes se distribuyen uniformemente a la hora de considerar qué tiene más peso práctico a la hora de considerar planes de carrera: 7 (50%) aseveran que se trata de la evaluación de desempeño y otros tantos opinan que es información consolidada en Metodologías Ágiles. La distribución de los respondientes sobre este apartado evidencia la relación con otros dos ítems anteriormente mencionados.

Aquellos respondientes que expresan mayor satisfacción con la herramienta y opinan que ella les da espacio y flexibilidad para definir objetivos y comportamientos, tienden a conferirle centralidad a la herramienta para este uso. Al contrario, quienes utilizan una herramienta con *feedback* anual o semestral (especialmente cuando el evaluado tiene menos espacio para dar su opinión y su evaluador padece una menor flexibilidad de la herramienta para consolidar sus observaciones) tienden a identificar a la dinámica cotidiana como la fuente principal de sus oportunidades de desarrollo.

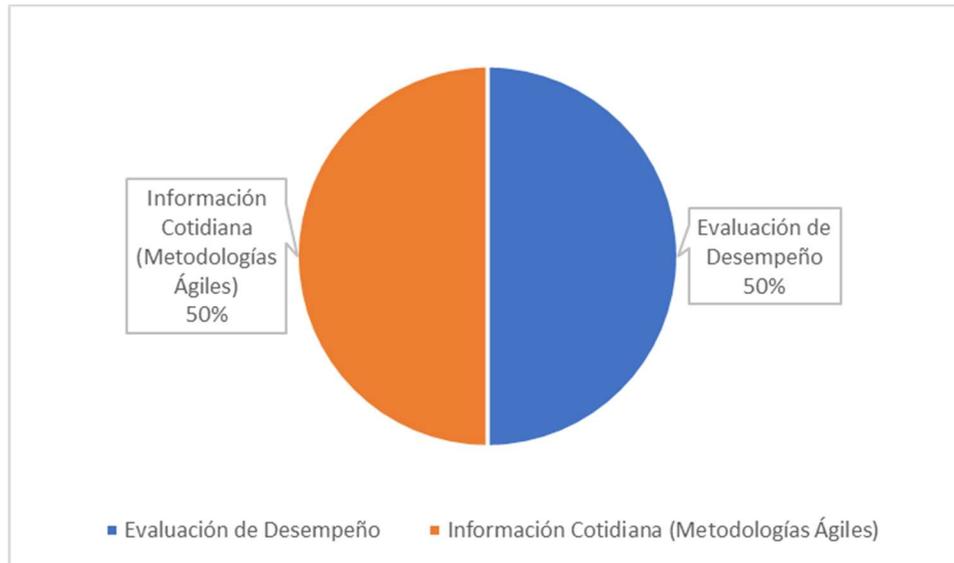


Figura 19 - Percepción de Mayor Peso Práctico (Planes de Carrera Individuales)

A lo largo de este capítulo, se ha caracterizado a los participantes de la muestra, relevado su percepción sobre sus Herramientas de Evaluación de Desempeño y enumerado las métricas que recaban sobre Metodologías Ágiles. En el siguiente Capítulo IV se enriquece esa información con los comentarios adicionales que cada respondiente proveyó a lo largo de la entrevista. Luego, se reestructuran esos datos en forma de propuestas que ayudan a sortear los desafíos más comunes y contribuyen a implementar sistemas que articulan exitosamente las Herramientas de Evaluación de Desempeño y las métricas provistas por la dinámica cotidiana en equipos que siguen Metodologías Ágiles de Desarrollo de Software. La formulación se basa, esencialmente, en eludir las prácticas que han afectado negativamente la percepción de ambos sistemas e implementar aquellas que contribuyen a que el funcionamiento de ambos sea considerado satisfactorio.

## Capítulo IV: Hacia una Implementación Satisfactoria

En el capítulo precedente se han descrito los métodos de evaluación y las métricas más populares actualmente en uso, las razones detrás de cada implementación y los beneficios y problemas que los respondientes han detallado para cada una de ellas. En este capítulo se condensa ese volumen de información subjetiva (agregando además comentarios y recomendaciones que han sido capturados en notas que trascendieron el cuestionario guía) y se lo convierte en recomendaciones. Estas recomendaciones procuran aumentar las chances de implementar Herramientas de Evaluación del Desempeño y una Metodología Ágil de desarrollo de software e integrarlas beneficiosamente.

### 1. Hacer foco en Objetivos por Área y Objetivos/Comportamientos Individuales

Como se habrá apreciado, las herramientas de evaluación de desempeño más tradicionales no están entre las preferidas de la mayoría de las empresas relevadas. Se refrenda, por lo tanto, la impresión de que el espíritu de las evaluaciones en empresas de servicios de desarrollo de software está migrando hacia la gestión de objetivos y comportamientos. Es deseable que así sea: bien implementados, esos métodos de gestión del desempeño consiguen adaptar los criterios de evaluación a cada rol y a cada individuo, resultando de ese modo más exigentes, más motivantes y mucho más susceptibles de identificar y desarrollar recursos con potencial.

La primera recomendación que se desprende del presente trabajo, entonces, es precisamente esa: si aún no lo hubieran hecho, es beneficioso crear un sistema de evaluación de desempeño basado en definir objetivos alcanzables específicos para cada área y comportamientos con observables relevantes para cada rol o individuo. La mayor eficacia de un sistema se presentará

cuando cada rol tiene sus indicadores específicos, y cada nivel de responsabilidad demanda un grado de cumplimiento para cada uno de los observables. Si la herramienta está bien diseñada, los indicadores y los niveles a alcanzar se desprenderán naturalmente desde la estrategia de la empresa. Por otra parte, la adaptabilidad a las características personales de cada evaluado puede manejarse como un segundo grupo de objetivos a medida o (tal como se propone) una serie de comportamientos.

El propósito último es que en la evaluación de desempeño de cada individuo se defina un criterio de éxito con relación directa a su aporte al colectivo, enriqueciendo posteriormente el análisis a través de la definición de áreas de trabajo centradas en las debilidades de cada evaluado. Desde este enfoque, será perfectamente posible que un evaluado tenga una cantidad de indicadores o planes de acción divergente de la de sus pares, y consecuentemente que sus evaluaciones no sean equiparables. Por ese motivo, en esos escenarios se sacrifica la posibilidad de efectuar comparaciones directas entre evaluados: es un costo que a largo plazo resulta mucho menos significativo que el provecho que aporta la herramienta desde la óptica de la motivación, la colaboración entre los equipos y el desarrollo personal.

## **2. Quitar a la calificación del centro de la escena**

La segunda recomendación (y seguramente la más ambiciosa) es desafiar la necesidad de resumir el ejercicio de evaluación de desempeño con una calificación general. Se acepta que el uso más extendido de la información provista por las evaluaciones continúa siendo determinar ajustes de salarios, como también que ese ejercicio es más fácil de realizar cuando hay una nota general que resuma el contenido. No obstante, como se postuló en el apartado teórico de este trabajo, esa función termina por adquirir demasiada centralidad en el ejercicio, dándole excesivo peso a la nota

final y difuminando el valor de alcanzar consensos sobre las áreas más sustanciosas: evaluación de objetivos, observación de comportamientos y acuerdos al respecto entre evaluador y evaluado.

### **3. Debilitar el lazo entre calificación y remuneración**

En aquellos esquemas donde la relación entre retribución individual y nota general de la evaluación están indisolublemente unidos, puede que sea mejor pensar en rediseñar el sistema de compensaciones para que cada nivel de rendimiento (es decir, la combinación entre rol, experiencia en el rol y capacidad técnica) tenga una retribución competitiva a nivel de mercado. El ejercicio ha probado ser exitoso en empresas del rubro y simplifica los cálculos sobre retribuciones individuales al mismo nivel que una normalización; por otro lado, ahorra burocracia y preserva las relaciones entre los equipos y sus líderes, dado que las instancias de negociaciones salariales individuales dejan de ser un factor negativo. El alcance de estas modificaciones sobre los esquemas de retribución es significativo; hay plena consciencia de los desafíos que una modificación así implica. No obstante, quitarle centralidad a ese aspecto tiene efectos beneficiosos en términos de planes de desarrollo y estrategia motivacional. Por eso, se puede dejar de mencionar la importancia de relativizar las notas generales de las evaluaciones de desempeño como una tercera recomendación.

### **4. Elegir las métricas más relevantes y usarlas para robustecer las evaluaciones**

La cuarta recomendación pasa por recolectar métricas, especialmente las que respectan a Metodologías Ágiles, pero también aquellas que requieren hacer un seguimiento de las tareas de *testers* y programadores. Además de la hibridación metodológica (que muchas veces los consultores se ven obligados a incorporar por demandas del proyecto o de los clientes), una de las amenazas más frecuentes es aquella que vacía de propósito a la Metodología Ágil. Por ejemplo,

creyendo que la centralidad está dada por la recurrencia y estilo de las reuniones diarias, el uso de etiquetas para cada tarea y demás aspectos que podríamos considerar rituales.

Es cierto que, como se menciona en el Capítulo II, la Metodología Ágil alienta a rechazar tareas que no contribuyan explícitamente a resolver las tareas incluidas en cada *Sprint*. Esto no significa en absoluto que las Metodologías Ágiles desprecien la documentación; antes bien, demandan que se la estructure en forma de *acceptance criteria* tal y como es expresado por el cliente, de modo simple y directo. Tampoco puede prescindir de las estimaciones; sólo puede optarse por estimar esfuerzo solamente para las tareas de más alta prioridad (que se implementarán antes) y dejar pendientes las estimaciones del trabajo que quedará por un tiempo en el *Backlog*. Hay argumentos como estos para cada una de las prácticas recomendadas: no perder excesivo tiempo en reuniones, medir el agregado de tareas que no estaban planificadas, detectar y superar errores de proceso, etcétera. Las entrevistas del capítulo III indagaron sobre el uso de métricas que dieran cuenta de la adherencia a cada una de esas prácticas recomendadas.

Por este motivo, las Metodologías Ágiles no tienen un modo práctico de ser exitosas prescindiendo totalmente de las métricas. A través de las herramientas de uso cotidiano es relativamente simple obtener un termómetro del progreso diario, utilizando al menos el análisis del *Burn-Down chart*, pero idealmente todas y cada una de las métricas mencionadas en el apartado de Metodologías Ágiles y del Negocio presentadas en los Capítulos II y III. Muchas de ellas se desprenden de modo simple partiendo del análisis de los tableros de trabajo y las reuniones de planificación. Incluso en implementaciones híbridas, con el paso del tiempo se alcanza un volumen de información que permite refinar estimaciones futuras, encontrar las causas de los desvíos, resolver problemas de proceso y mitigar los daños en la relación con el cliente cuando un compromiso se ve amenazado.

Puede que cada equipo de desarrollo necesite libertad para interpretar cada una de las métricas, especialmente porque conocerán mejor que nadie el proceso real que se sigue desde que se recibe un requerimiento hasta que se termina de implementarlo. Se puede aceptar que cada líder tenga un gusto particular por ciertas métricas, e incluso que los problemas más relevantes o recurrentes sean más fácilmente manejables prestando atención a unas sobre otras. Lo ideal, en este contexto, es que la estrategia de la empresa en lo que refiere a los equipos de desarrollo sea dedicar el esfuerzo a recolectar las métricas, y exigir un grado mínimo de análisis y operacionalización sobre los indicadores que cada líder fundamente como centrales. La discrecionalidad y criterios divergentes bien pueden aceptarse, tolerarse o alentarse (dependiendo del caso) en esta última etapa.

Como mínimo, se recomienda que se profundice el análisis a nivel grupal e incluso individual para los casos extremos: aquellos que arrojan estadísticas que parecen inusualmente productivas y de alta calidad (para identificar recursos de alto potencial, mejoras en los procesos o incluso premiar especialmente a un equipo por un trabajo exitoso) o, por el contrario, estimaciones llamativamente imprecisas y productos de baja calidad o con defectos de diseño (para detectar fallos en el proceso y proponer mejoras, detectar necesidades de supervisión, cursos o apoyo a los equipos e individuos que no lograron completar una etapa de desarrollo en tiempo y forma). En otras palabras: las estadísticas y métricas al menos tienen que caracterizar todo el trabajo de la empresa y evidenciar *outliers* (los casos atípicos, lo inusual) sobre los cuales es necesario profundizar el análisis para reconocer lo bueno y corregir lo defectuoso. Siendo que un gran volumen de datos debería estar disponible de todas maneras, es un gran beneficio que acarrea relativamente poco esfuerzo adicional.

En lo que respecta a métricas sobre desarrollo y calidad, pueden también quedar a criterio de cada líder de equipo de desarrollo, en tanto se capturen algunas.

Del primer grupo, se sostiene que los *Peer-Review Ratings* son la métrica más recomendable. Si bien su confección es casi artesanal, su interpretación frecuentemente es subjetiva y es la práctica que mayor esfuerzo requiere para producirse, también es la más descriptiva y la que redundará en mayores beneficios a largo plazo: de ella se desprenderán una mejor calidad de código, un menor esfuerzo de mantenimiento y (en casos donde eso es viable) una deuda técnica más manejable. Como ventaja extra, si es llevada a cabo por los expertos de cada uno de los equipos es virtualmente imposible de falsear (hay que tener presente que cuando las demás métricas se comunican, los programadores podrán buscar formas de cambiar sus prácticas para mejorar los resultados artificialmente) y no demanda más tiempo de análisis: el propio ejercicio se cierra con conclusiones y cursos de acción.

Del apartado de Calidad de Software pueden también elegirse las métricas que controlen mejor las partes del proceso que intuitivamente se noten más débiles (o bien, las que tengan más impacto para el negocio). Es una buena idea empezar por las más elegidas según se detalla en el Capítulo III: aquellas que vinculan la calidad del código con defectos encontrados durante las pruebas de aceptación del usuario, fallos encontrados por usuarios finales en sitios o versiones ya hechas públicas y cantidad de defectos encontrados por *Story/Task*. Todas ellas tienen la enorme ventaja de ser más fácilmente recopiladas que las métricas de desarrollo referidas en el párrafo anterior, aunque el mejor provecho (especialmente para no implementar las correcciones cuando es demasiado tarde) se obtiene implementando métricas a todo nivel: metodológico, de desarrollo y de calidad. Con esto se redondea la cuarta propuesta.

## 5. Las métricas no reemplazan a la información subjetiva

La quinta recomendación es resistir a la tentación de apoyarse solamente en las métricas para las evaluaciones de desempeño, lo cual constituiría una limitación igual de severa que no recopilar datos en absoluto. El análisis debe estar apoyado sobre métricas e indicadores, pero desde el momento en que ellas requieren de análisis e interpretación ya estaremos incluyendo una porción significativa de subjetividad por parte del evaluador o quienes estén a cargo de interpretar los datos. ¿Cuánta más complejidad podría agregarle a ese aspecto el incluir observaciones de pares y líderes e integrarlas a cada evaluación?

Si bien los respondientes han especificado que en buena medida existe *feedback* subjetivo y se lo incorpora a la evaluación (ya sea como notas, planes de acción sobre comportamientos u objetivos individuales como participar de talleres de comunicación o gestión del tiempo), también han dejado claro que continúa siendo mejorable la integración de los aspectos subjetivos y de procesos, especialmente en lo que respecta a Metodologías Ágiles. Es unánime la opinión de que las reuniones retrospectivas arrojan información valiosa a nivel individual y grupal; también ha sido unánime la opinión de que el trato y seguimiento de esas recomendaciones es informal, desordenado y depende de la voluntad de cada líder. Así, suelen pasar meses desde que el grupo produce recomendaciones de cambio y que esas opiniones se traducen (cuando es necesario) en reconocimientos o planes de acción dentro de las evaluaciones de desempeño. El tiempo suele conspirar contra las percepciones de importancia que dichos aspectos tuvieron al momento de ser mencionados, y hasta puede empujar al olvido aspectos centrales en casos extremos donde ni siquiera se documenta formalmente aquello que ha discutido grupalmente.

Prácticas como *Feedback 360* aplicadas una vez al año no son un buen sustituto de la consolidación sistemática de información subjetiva o cualitativa. Más allá de la objetividad adicional que aporta consolidar opiniones de todos aquellos que se relacionan con un evaluado particular, la técnica no está exenta de las limitaciones subjetivas mencionadas en el Capítulo I (como el efecto halo, miopía y afinidad). Tampoco genera beneficios adicionales respecto a la erosión de la información producto del tiempo transcurrido; es probable que esos elementos sean olvidados por cada participante del mismo modo en que han sido olvidados por algunos miembros del grupo. Por eso, celebrar reuniones retrospectivas, documentar las lecciones aprendidas y especialmente las recomendaciones grupales (y su alcance individual cuando lo hubiera) son aspectos donde la mayoría de los respondientes coinciden en que hay oportunidades de mejora. Alimentar las recomendaciones sobre aspectos subjetivos y cualitativos y reforzar estas impresiones con observables recolectados en esas reuniones es, en suma, otra propuesta para enriquecer las evaluaciones de desempeño.

## **6. Involucrar a la dirección, capacitar a los líderes, supervisar el proceso**

La última recomendación es la de no subestimar la importancia del entrenamiento y supervisión de los líderes, especialmente cuando la dirección se involucra en forma directa. Las entrevistas de características como las propuestas (basadas en objetivos y comportamientos, adaptadas a cada evaluado, con respaldo en observables, métricas y aspectos subjetivos) son una herramienta muy poderosa, con alta percepción de justicia y relevancia... si el evaluador/líder se las toma con la máxima seriedad. Se ha visto en el Capítulo III que los ejercicios de calibración no están muy extendidos, siendo alta la cantidad de respondientes que consideran que no existe en absoluto y una mayoría de casos en los que se basan en la opinión del jefe del jefe/evaluador. Este

último enfoque no reduce demasiado la dependencia de aspectos subjetivos, es decir, de la voluntad y aplicación de cada uno de los involucrados en la cadena de evaluación.

El ejercicio será más provechoso si los lineamientos para evaluación se pueden traducir a una calibración sobre *role models* presentes y pasados para cada equipo. La ventaja de usar como patrón los comportamientos de miembros destacados de un grupo es que es fácil convertir comportamientos en observables, hay escasas dudas sobre la posibilidad de alcanzarlos y dichos criterios han tenido un impacto evidentemente positivo en el negocio. Puede también ser eficaz dividir esos patrones en categorías que se desprenden de la cultura de la empresa, con observables para cada una que reflejan lo que hicieron otros *role models* en el pasado; así, además de su utilidad con fines de calibración, serán también un propagador de los valores de la compañía.

Finalmente, se reitera un aspecto mencionado en el Capítulo I, estrechamente vinculado con esta sexta recomendación. Para que los líderes se dediquen con la máxima seriedad al ejercicio de evaluación, deben contar con el tiempo necesario para hacerlo y se debe incluir la calidad de las evaluaciones de desempeño como un elemento a integrar en sus propias evaluaciones. De poco servirá preparar a los líderes para realizar un ejercicio al que luego se lo posterga o degrada frente a urgencias cotidianas o (peor aún) cuyas consecuencias nadie analiza.

Por este motivo, la dirección es el único garante de que la herramienta de evaluación cuente con la capacitación, la disponibilidad y la evaluación sobre los avances en los ejercicios, que son los tres pilares que asegurarán que la eficacia de la implementación sea al menos aceptable y se minimice la dispersión en la calidad de las evaluaciones de un líder a otro. Si esos tres pilares están presentes y son implementados de modo saludable, el ejercicio tendrá muchas más chances de ser provechoso.

## Conclusiones

- Durante los últimos años, ha comenzado un proceso de transformación de las evaluaciones de desempeño para empresas de servicio o de producto que tienen equipos de desarrollo de software. En los estudios de caso no se ha llegado a percibir una mayor tendencia a abandonar todo ejercicio de evaluación, pero sí se está avanzando hacia herramientas menos estructuradas. El *Management* a través de objetivos grupales, objetivos individuales y comportamientos parece estar desplazando a herramientas como distribuciones forzadas y normalizaciones, efecto razonable si se tienen en cuenta las limitaciones de estas últimas y los costos que aparejan en términos de motivación, rendimiento y vínculos con pares y superiores.
- Esta agradable tendencia todavía tiene un largo camino por recorrer. El cuerpo práctico del presente trabajo todavía evidencia que los comportamientos y objetivos verdaderamente valiosos no siempre están siendo correctamente relevados. En muchos casos, eso obedece a que la herramienta de la empresa busca ser suficientemente genérica como para que sea apta para ser implementada sobre cualquier rol. En otros, las limitaciones ocurren porque la flexibilidad de la herramienta se convierte en un arma de doble filo: la dependencia de la buena voluntad, ecuanimidad y esmero por parte del evaluador hace que los resultados del ejercicio sean muy diferentes según quién lo lleve adelante.
- Se ha propuesto abordar estos desafíos con las recomendaciones de evaluar si es imprescindible sostener la existencia de calificaciones generales, la de definir objetivos y comportamientos individuales (incluso si eso requiere sacrificar la posibilidad de comparar el rendimiento entre evaluados) y la de reforzar la capacitación, disponibilidad y

seguimiento de los evaluadores. Los evaluados y la empresa en pleno lo agradecerán: aquellos que han refinado de modos similares sus actuales herramientas de desempeño se manifiestan conformes con sus aportes, aseveran que lo que ellas miden es relevante para la estrategia y rentabilidad de la empresa, la perciben como una aliada para el desarrollo personal y la entienden como un paso más hacia la adopción de la cultura de la empresa.

- Se ha confirmado la expansión y el alcance de las Metodologías Ágiles de desarrollo de software. Con mayor o menor grado de pureza e hibridación, todas las empresas relevadas las han adoptado. Estas implementaciones no han estado exentas de tensiones con procesos internos o aquellos que impactan en los de sus clientes. Quizás no sea reprochable hacer concesiones menores para poder participar en un proyecto; sí merecen replanteos aquellas implementaciones que sacrifican la esencia misma de las metodologías. Quienes optan por sacrificar las estimaciones, estructurar el trabajo pendiente a partir de documentos kilométricos en vez de descomponerlo en *User Stories*, fijar duraciones de ciclos demasiado hipertrofiadas y (especialmente) no medir los avances ni celebrar reuniones retrospectivas, terminan descubriendo que se ve muy mermada su capacidad de implementar soluciones en tiempo y forma. Ser inflexible y purista para con la metodología puede sonar como una demanda exagerada, pero la práctica indica que el extremo opuesto trae aún más disgustos.
- Por los mismos motivos, se recomienda que quienes efectivamente siguen Metodologías Ágiles de desarrollo utilicen la valiosa información que la propia dinámica cotidiana genera. Como mínimo, se debe analizar diariamente el volumen de trabajo pendiente, los desvíos en las estimaciones, el cumplimiento de previsiones de fechas y costos y la satisfacción del cliente. Si se siguen las prácticas esenciales, la misma herramienta de

gestión de proyecto ofrece esta información y solamente demanda bajo esfuerzo para consolidarla, interpretarla, crear accionables, documentarlos y supervisar los progresos.

- No obstante, desde una perspectiva de desarrollo de software no se puede dejar librada la evaluación del trabajo del equipo solamente a ese nivel. La tarea de los programadores será exitosa sólo si además de cumplir con objetivos inmediatos produce código escalable, mantenible, sencillo de interpretar y con la mínima deuda técnica posible. Por lo tanto, al menos es necesario consensuar criterios y celebrar ejercicios periódicos de *peer-code reviews*. Esta práctica saludable (e imprescindible) debe idealmente reforzarse con otras métricas relevantes para cada equipo y sus circunstancias. Es pertinente elegir ciertas métricas por sobre otras; por ejemplo, empresas de producto pueden darle mucha trascendencia a la disponibilidad del sitio productivo y a la deuda técnica mientras que empresas que tercerizan servicios pueden prescindir de ellas (cuando no son responsables últimas de las implementaciones) para enfocarse en la velocidad del equipo. Lo relevante es que esas métricas se capturen: las herramientas de gestión y los repositorios de código lo vuelven relativamente simple.
- Desde el punto de vista de la calidad también hay espacio para que cada equipo seleccione las métricas que le resulten más relevantes. Es recomendable empezar haciendo foco en las tres métricas más populares entre las relevadas en el cuerpo práctico: la proporción de defectos encontrados durante las pruebas de aceptación del usuario, la existencia (o no) de fallos encontrados por usuarios finales en aplicaciones en faz productiva y la cantidad de defectos encontrados por *Story/Task*. Todas ellas son simples de capturar, rápidas de interpretar y tienen la ventaja de convertirse en recomendaciones a nivel grupal e individual con la misma sencillez.

- Es tan nocivo reposar solamente en criterios subjetivos como conferir toda la centralidad a la recolección e interpretación de las métricas. El modo más sencillo de buscar un balance es identificar los casos que más se desvían del promedio o de las expectativas, para entonces intensificar el análisis y buscar las causas raíz solamente para esas ocurrencias puntuales. Una vez corregidos los problemas, robustecido el proceso y reconocido el rendimiento donde fuera pertinente, se puede repetir el ejercicio (nuevamente, haciendo foco en los casos que más se desvían de las expectativas). A medida que el ejercicio y el equipo de trabajo maduren, los ajustes se volverán más sutiles y la mejora será evidente tanto a nivel subjetivo como a partir del análisis de las métricas.

Las empresas relevadas que han seguido estas recomendaciones a la hora de medir el desempeño han pasado a disponer de herramientas para gestionar a sus equipos, incentivar el desarrollo de sus profesionales, promover a los recursos de mayor potencial, estimular y retener a sus talentos, aumentar la satisfacción de sus clientes, detectar y rectificar errores en sus procesos, mejorar las capacidades de su personal y representar mucho más eficazmente los valores de la empresa a través del comportamiento de sus empleados.

Es un trabajo arduo que debe ser hecho artesanalmente y casi a medida, como único camino para maximizar su eficacia. Las observaciones vertidas en el presente trabajo sugieren que el esfuerzo genera un gran retorno.

## Bibliografía

Alnaji, L. & Salameh, H. (2015). *Performance-Measurement Framework to Evaluate Software*

*Engineers for Agile Software-Development Methodology*. Obtenido de:

<https://www.iiste.org/Journals/index.php/EJBM/article/view/19143/19842>

Fecha de Consulta: 28/03/2019.

Beck, K., Beedle, M., van Bennekum, A. & Otros (2001). *Twelve Principles Behind the Agile*

*Manifesto*. Obtenido de: <https://agilemanifesto.org/principles.html>

Fecha de Consulta: 28/03/2019.

Berzisa, S. & Podjavo, I. (2017). *Performance Evaluation of Software Development Project*

*Team*. Obtenido de: <http://journals.ru.lv/index.php/ETR/article/download/2543/2562>

Fecha de Consulta: 28/03/2019.

Casanova, L. (2016). *Evaluaciones de Desempeño: ¿una especie en extinción?* Mimeo.

Cohn, M. (2006). *Agile estimating and planning*. Upper Saddle River, NJ: Pearson Education.

Dikert, K., Lassenius, C. & Paasivaara, M. (2016). *Challenges and success factors for large-*

*scale agile transformations: a systematic literature review*. Obtenido de:

<https://www.sciencedirect.com/science/article/pii/S0164121216300826>

Fecha de Consulta: 28/03/2019.

Gonçalvez, L. (2019a). *What is Agile Methodology*. Entrada en blog personal. Disponible en:

<https://luis-goncalves.com/what-is-agile-methodology/>

Fecha de Consulta: 28/03/2019.

Gonçalves, L. (2019b). *Burndown Chart – The Ultimate Guide for Every Scrum Master*. Entrada en blog personal. Disponible en: <https://luis-goncalves.com/burndown-chart-ultimate-guide/>

Fecha de Consulta: 28/03/2019.

Hatum, A. (2009). *Alineando la organización: Estrategia y prácticas de recursos humanos para managers*. Buenos Aires: Granica.

Hatum, A. (2011). *El futuro del talento: Gestión del Talento para Sobrevivir la Crisis*. Buenos Aires: Temas.

Inman, M. (2010). How a Web Design Goes Straight to Hell. Entrada en blog comercial.

Obtenido de:

[https://theoatmeal.com/comics/design\\_hell](https://theoatmeal.com/comics/design_hell) Fecha de Consulta: 28/03/2019.

Klein, M., Bitzer, S. & Schumann, M. (2012). Towards IT-Supported Management by Objectives – A Design Theory Based on Intra-Organizational Weblogs. *ECIS 2012 Proceedings, #14*.

Obtenido de <http://aisel.aisnet.org/ecis2012/14> Fecha de Consulta: 28/03/2019.

Hodges, N. (2012). *Measuring Developer Productivity*. Disponible en:

<http://www.nickhodges.com/post/Measuring-Developer-Productivity.aspx>

Fecha de Consulta: 28/03/2019.

Holland, K. (2006). Performance reviews: Many need improvement. *New York Times*, edición del 10 de septiembre de 2006. Obtenido de:

[http://www.nytimes.com/2006/09/10/business/yourmoney/10mgmt.html?\\_r=0](http://www.nytimes.com/2006/09/10/business/yourmoney/10mgmt.html?_r=0)

Fecha de Consulta: 28/03/2019.

Lowe, S. (2018). *9 metrics that can make a difference to today's software development teams*.

Obtenido de: <https://techbeacon.com/app-dev-testing/9-metrics-can-make-difference-todays-software-development-teams> Fecha de Consulta: 28/03/2019.

McHugh, S. (2012). *How to Not Destroy your Agile Team with Metrics*. Obtenida de:

<https://www.infoq.com/articles/not-destroy-team-metrics>. Fecha de Consulta: 28/03/2019.

Mercer (2013). *2013 Global Performance Management Survey Report*. Disponible en:

<https://www.mercer.com/content/dam/mercer/attachments/global/Talent/Assess-BrochurePerfMgmt.pdf>. Fecha de Consulta: 28/03/2019.

Molinari, P. (2017). *¡Basta de la vieja gestión del desempeño!* Entrada en blog personal.

Obtenido de: <http://www.pmolinari.com/2017/04/13/basta-de-la-vieja-gestion-del-desempeno/>

Fecha de Consulta: 28/03/2019.

Montano, C. (2019). *Accelerate Delivery with Quality: What the Automotive Industry can teach*

*us about QA*. Obtenido de: <https://fortegrp.com/blog/software-development/accelerate-with-quality/>

Fecha de Consulta: 28/03/2019.

Pérez, A. (2017). *4 Técnicas para estimar: PERT, Delphi, Planning Poker, T-shirt*. Entrada en

blog personal. Obtenida de: <http://www.ceolevel.com/4-tecnicas-para-estimar-pert-delphi-planning-poker-tshirt>

Fecha de Consulta: 28/03/2019.

Pierce, M. (2011). *The annual performance review sucks; try an agile appraisal!* Entrada en blog

personal. Obtenida de:

<https://blog.mikepearce.net/2011/04/14/the-annual-performance-review-sucks-try-an-agile-appraisal/> Fecha de Consulta: 28/03/2019.

Poppendieck, M. (2004). *Unjust Desserts*. Obtenida de:

<http://www.poppendieck.com/pdfs/Compensation.pdf>. Fecha de Consulta: 28/03/2019.

Provenzano, G. (2018). *Individual Productivity : How Do You Measure It in Your Team?* Obtenida

de: <https://www.knowledgehut.com/blog/project-management/individual-productivity-how-do-you-measure-it-in-your-team>. Fecha de Consulta: 28/03/2019.

Romeu, A. (2014). *Planning poker – ¿Cómo planificar un sprint en Scrum?* Entrada en blog personal. Obtenida de: <https://albertoromeu.com/scrum-planning-poker/> Fecha de Consulta: 28/03/2019.

Schwaber, K. (2004). *Agile Project Management with Scrum*. USA:Microsoft Press.

Smith, G., & Sidky, A. (2009). *Becoming agile ... in an imperfect world*. Greenwich, CT: Manning.

Spolsky, J. (2000). *Incentive Pay Considered Harmful*. Disponible en:

<https://www.joelonsoftware.com/2000/04/03/incentive-pay-considered-harmful/>

Fecha de Consulta: 28/03/2019.

Spolsky, J. (2009). *Why I Never Let Employees Negotiate a Raise*. Disponible en:

<https://www.joelonsoftware.com/2009/04/01/fog-creek-compensation-system/>

Fecha de Consulta: 28/03/2019.

Sutherland, J. (2018). *Performance Appraisals*. Disponible en:

<https://www.scruminc.com/performance-appraisals/> Fecha de Consulta: 28/03/2019.

York, B. (2015). *The Best Developer Performance Metrics*. Disponible en:

<https://medium.com/@yupyork/the-best-developer-performance-metrics-6295ea8d87c0>

Fecha de Consulta: 28/03/2019.

## Apéndices / Anexos

### Apéndice I: Cuestionario Guía para las Entrevistas

1. Dirección de Contacto

2. ¿Cuál es su nombre? ¿Qué rol desempeña en la empresa para la que trabaja? ¿Cuál es el nombre de la compañía que representa?

3. Por favor, marque el/los roles que ha desempeñado en su posición actual. [Elección Múltiple]

Evaluador;  
Evaluado;  
Procesador de Información;  
Receptor de la Información Consolidada;  
Otros (Especificar Cuál):

4. ¿Cuántos Equipos de Desarrollo de Software hay en su empresa (aproximadamente)?

5. ¿Qué tamaño tiene un equipo Ágil que usted evalúe como típico de la organización?

6. ¿Cuántos desarrolladores trabajan en ese equipo "típico"?

7. ¿Tiene su empresa un sistema de evaluación de desempeño? (Sólo 1 Opción)

Sí;  
No, nunca tuvo;  
No, lo tuvo en el pasado y fue abandonado.

### Herramientas de Evaluación de Desempeño

8. ¿Con qué frecuencia se dan las reuniones individuales de Feedback? (Sólo 1 Opción)

5+ veces por año;  
Cada 3 meses, aproximadamente;  
Entre 3 y 2 veces por año;  
De frecuencia aproximadamente anual o menos.

9. ¿Cuáles de las siguientes métricas/herramientas individuales son capturadas o consolidadas por el sistema implementado actualmente? [Elección Múltiple].

- Calificación General o Total;
- Objetivos Individuales;
- Comportamientos/Competencias Individuales;
- Balanced Scorecard;
- Ponderación de Objetivos o Comportamientos (uno a uno);
- Ponderación de Objetivos o Comportamientos (por tipo);
- Objetivos por Área (que se desprenden de la estrategia);
- Objetivos por Área (que cascadean informalmente de la estrategia);
- Evaluación de Habilidades de Manejo de Equipo;
- Evaluación de Habilidades de Manejo de Equipo (que impactan en la remuneración);
- Ninguna de las Anteriores.

10. ¿Qué métricas no mencionadas en el apartado anterior forman parte del actual sistema de medición de desempeño individual?

11. ¿Cuáles de las siguientes metodologías forman parte de su sistema de evaluación de desempeño individual? [Elección Múltiple].

- Calibración de Objetivos (para roles similares);
- Calibración de Calificaciones;
- Autoevaluaciones;
- Evaluación 360 (o múltiple) – Formal;
- Evaluación 360 (o múltiple) – Informal;
- 9-box;
- No Sabe / No Contesta;
- Ninguna de las anteriores;
- Otros (especificar).

12. Para asegurar que haya distintos resultados de las evaluaciones de desempeño dentro de un mismo equipo, ¿cuáles de las siguientes herramientas forman parte del actual sistema de evaluación de desempeño? [Elección Múltiple].

- Distribución Forzada de Calificaciones;
- Lineamientos sobre la Distribución Esperada de las Calificaciones;
- Ranqueo Relativo de Individuos (dentro de un equipo);
- Ranqueo Relativo de Individuos (entre varios equipos);
- Ninguna de las anteriores.
- Otros (especificar):

13. ¿Cuáles de los siguientes modos de calibración forman parte del actual sistema de evaluación de desempeño? [Elección Múltiple].

- Calibración informal;
- Calibración forzosa;
- Revisión del jefe del jefe/Evaluador;
- Ninguna de las anteriores;
- Otros (especificar).

14. ¿Cuáles de las siguientes acciones toman información provista por el actual sistema de gestión individual del desempeño? [Elección Múltiple].

- Determinar ajustes salariales por performance;
- Determinar ajustes salariales por potencial;
- Determinar pagos únicos (bonos) por performance;
- Determinar pagos únicos (bonos) por potencial;
- Vincular calificaciones con planes de desarrollo;
- Crear planes de desarrollo individuales;
- Implementar planes de mejora de desempeño;
- Vincular calificaciones con promociones;
- Vincular calificaciones con decisiones de sucesión;

Ninguna de las anteriores;

Otras (especificar):

15. ¿En cuál o cuáles de las siguientes acciones del programa de evaluación de desempeño participan la dirección y/o la gerencia? [Elección Múltiple].

Discusiones de Planificación de Objetivos;

Discusiones de Performance Individuales (cara a cara);

Discusiones de Performance Grupal (cara a cara con el equipo);

Dando coaching/feedback de modo regular;

Conversando con la dirección sobre la Gestión del Desempeño;

Ninguna de las anteriores;

Otras (especificar).

### **Metodologías Ágiles**

¿Cuáles de las siguientes métricas son capturadas o consideradas relevantes por equipos operativos de la organización?

16. Relacionadas con la Metodología y el Negocio - Sólo Grupales [Elección Múltiple].

Análisis del Burn-Down Chart (o Gráficos de Trabajo Pendiente);

Tiempo promedio para completar una nueva feature;

Cumplimiento de Schedule Forecast/Cost Forecast;

Tiempo de Ciclo (desde que se inicia una task hasta que se la promueve a PROD);

Lead Time (entre que una task entra al *backlog* y se la promueve a PROD);

ROI del Cliente;

Aumento de Rentabilidad del Cliente.

17. Relacionadas con la Metodología y el Negocio - Individuales y/o Grupales. [Elección Múltiple].

Tiempo promedio entre estados “código terminado” y “tarea completa”;

Velocidad [cantidad de *story points* u horas completadas por *Sprint*];

Commit-to-deploy time

Período/Frecuencia con que se promueven nuevos cambios/features a producción

Satisfacción del Cliente (subjetiva - feedback post Sprint);

Horas/Sprint Invertidas en *Meetings*;

Desvío entre horas estimadas y efectivamente invertidas en una Task/Subtask;

18. Relacionadas con Web/App Development [Elección Múltiple – Grupal/Individual].

Peer Code Review Rating

# Lines of Code

# Commits/Check-Ins

# Pull Requests Aceptados

Eficiencia del Código (%)

Tasa de Código Improductivo (%)

Uptime/Downtime

Code Flexibility

Technical Debt

19. Relacionadas con la Calidad del Código [Elección Múltiple – Grupal/Individual].

Defectos por Story/Task;

Vida Promedio de los Defectos;

Tasa de Rechazo/Task;

Cantidad de Casos de Prueba Aprobados;

% Tiempo Total Invertido en Defectos;

Defectos UAT / Totales;

Cantidad de Defectos en PROD.

20. ¿Cómo se consolida el *feedback* de las *retrospectives*? ¿Individual o Grupal? ¿Cómo se extiende ese feedback a las evaluaciones de desempeño? [Texto Libre]

### Percepción General del Sistema

21. ¿Qué grado de eficiencia cree que tiene el actual Sistema de Evaluación de Desempeño para medir los objetivos individuales? [Escala de Likert – Elegir 1 valor]

1 2 3 4 5

22. ¿Qué grado de eficiencia cree que tiene el actual Sistema de Evaluación de Desempeño para medir el potencial individual? [Escala de Likert – Elegir 1 valor]

1 2 3 4 5

23. ¿Qué grado de eficiencia cree que tiene el actual Sistema de Evaluación de Desempeño para medir la colaboración de cada individuo para con su grupo de trabajo? [Escala de Likert – Elegir 1 valor].

1 2 3 4 5

### Peso Práctico de cada herramienta

24. A la hora de considerar recategorizaciones individuales, ¿qué tiene más peso práctico?

Evaluación de Desempeño

Información cotidiana recolectada a través de prácticas alentadas por Metodologías Ágiles

25. A la hora de considerar retribuciones salariales individuales, ¿qué tiene más peso práctico?

Evaluación de Desempeño

Información cotidiana recolectada a través de prácticas alentadas por Metodologías Ágiles

26. A la hora de considerar planes de carrera individuales, ¿qué tiene más peso práctico?

Evaluación de Desempeño

Información cotidiana recolectada a través de prácticas alentadas por Metodologías Ágiles